JWST spies galaxies blazing in the infant universe p. 1280

Policy protections for kids in cyberspace p. 1294

Pumping fluids using fibers p. 1327

Science

$15
31 MARCH 2023
science.org
AAAS



AN ENDURING PARTNERSHIP

The long history between horses and Indigenous Americans pp. 1288 & 1316

VISIT…

LANZAROTE Caliente.com

EDITORIAL

Fix the backlash against public health

Societies generally have reacted to deadly epidemics by strengthening health systems, including laws. Under American federalism (the constitutional division of power between states and the federal government), individual states hold primary public health powers. State legislatures have historically granted health officials wide-ranging authority. After the anthrax attacks in the United States in 2001, the US Centers for Disease Control and Prevention (CDC) supported the Model State Emergency Health Powers Act, which granted public health officials even more expansive powers to declare a health emergency and respond swiftly. But all that ended with COVID-19, as state legislatures and courts gutted this authority. The next pandemic could be far deadlier than COVID-19, but when the public looks to federal and state governments to protect them, they may find that health officials have their hands tied behind their backs.

During the pandemic, over 30 states passed laws curbing health measures such as mask and vaccine mandates, quarantines, and business closures. Many state law reforms now allow the legislature to rescind executive health orders, effectively shifting decisions from infectious disease experts to lawmakers ill-equipped to make rapid, science-based judgments. The pandemic also unleashed more than 1000 lawsuits challenging COVID-19 measures put into place by health officials, a quarter of which were successful in invalidating them. The Supreme Court, with a conservative super-majority, struck down state COVID-19 restrictions on religious gatherings as well as the CDC's tenant eviction moratorium and vaccine-or-test rules for large employers put forth by the Department of Labor's Occupational Safety and Health Administration.

> "...state legislatures and courts have cut public health powers too deeply..."

The powers of health officials must strike a balance between health security and key societal values and constitutional rights, including personal freedom, economic activity, and educational opportunity. Indeed, health officials did overreach—for example, by closing schools and businesses for extended periods. But state legislatures and courts have cut public health powers too deeply, seriously jeopardizing future emergency responses. The legislative and judicial retrenchment during the pandemic signals the need for legal modernization. Instead of wholesale weakening of public health powers, states should modernize laws to balance powers and rights more productively, such as meaningful legislative checks on executive power exercised by health officials during crises, stronger criteria for executive orders concerning public health, and mechanisms for public input. But changes need to be far more surgical. Emergency powers for health officials were conceptualized for fast-moving situations, but COVID-19 broke the mold, with long-lasting deprivations of liberty.

Executive health officials should have strong and flexible authority to respond to emergencies, but legislatures should be empowered to modify or end a declaration of "emergency" that goes on too long or tramples personal freedom too far. Emergency powers of health officials also should include enhanced transparency and accountability. This might mean independent reviews of the exercise of emergency powers. Furthermore, health officials should evaluate the effects of interventions on disadvantaged populations, such as racial and ethnic minorities, lower-income families, migrants, and persons in congregate settings. Health equity should become a prevailing value. Health officials should engage with stakeholders and community leaders to mitigate unintended harms from emergency measures. Cross-agency collaboration during extended emergencies is also warranted, as education, commerce, transportation, and other sectors are increasingly affected.

Although states should remain the primary holders of emergency powers, expanding federal powers would provide a stronger baseline of national protections. Alongside updating the federal Public Health Service Act, as recommended by the National Academies, Congress could make emergency funding contingent on a state's compliance with national plans, providing a more unified national response and filling gaps in weak state responses. Modernizing federal public health authorities would also reduce litigation and judicial challenges. A National Commission on COVID-19 could begin a complex yet careful process to redesign legal systems and rebuild trust in public health authority. Future generations deserve well-calibrated laws that empower health officials to act decisively while mindfully safeguarding constitutional rights and social values.

–Lawrence O. Gostin and Sarah Wetter

Lawrence O. Gostin is University Professor at Georgetown University, where he directs the O'Neill Institute for National and Global Health Law, Washington, DC, USA. gostin@georgetown.edu

Sarah Wetter is an associate at the O'Neill Institute for National and Global Health Law at Georgetown University, Washington, DC, USA. sw1203@georgetown.edu

10.1126/science.adh9594

NEWS

"This has to fold into the way we manage other diseases. … If not now, then when?"

University of California, San Francisco, chair of medicine Bob Wachter, in *The Washington Post*, on reports that the White House plans to disband its COVID-19 response team in May.



IN BRIEF

Edited by **Jeffrey Brainard**

The ATLAS detector, which is fed by Europe's Large Hadron Collider, is as big as a cathedral.

PARTICLE PHYSICS

Particle mass dispels hint of new physics

A fleeting, weighty elementary particle called the W boson has just the mass predicted by theory, physicists working with Europe's Large Hadron Collider (LHC) reported this week at a conference in Italy. The finding comes from ATLAS, one of four large particle detectors fed by the LHC, and it contradicts the eyebrow-raising measurement reported last year in *Science* (8 April 2022, p. 125) that suggested the W was heavier than predicted by physicists' prevailing standard model. That discrepancy could have signaled that massive new particles lurking in the vacuum alter the mass of the W, which conveys the weak nuclear force just as the photon conveys electromagnetism's force. The hint that the W boson might be extra-hefty came from an analysis of data from a particle detector called D0, which was fed by the now-defunct Tevatron collider at the Fermi National Accelerator Laboratory. The analysis disagreed not only with the standard model, but also with D0's previous measurement. Now it is even more of an outlier.

NSF to survey LGBTQ identity

SCIENTIFIC COMMUNITY | The first tally of U.S. Ph.D. recipients who are LGBTQ may be available in 2026 if a pilot study announced last week by the National Science Foundation (NSF) is successful. Starting in July, NSF will test a series of questions about sexual orientation and gender identity on its Survey of Earned Doctorates, an annual census of the roughly 57,000 graduating doctoral students. Currently, the survey only offers a binary male-female choice for gender. NSF will use responses from the first year to determine which questions best capture information about LGBTQ students. The agency hopes to add a finalized set of questions to the survey starting in 2024 and to release data on sexual and gender identities in 2026. Advocates hope the information can be used to track whether LGBTQ people are underrepresented in science, technology, engineering, and math.

Nearby rocky planet looks bleak

ASTRONOMY | Researchers have used the new JWST orbiting telescope to take the temperature of a roughly Earth-size exoplanet—the first time observers have captured light from such a small rocky world. The nearby star TRAPPIST-1 is a major target for JWST because it is orbited by seven rocky planets—several potentially habitable. The star is also tiny and dim, making it easier to see the faint light of its planets. As expected, the planet closest to the star, TRAPPIST-1 b, is far from habitable, astronomers report this week in *Nature*: It has a surface temperature of 230°C and no detectable atmosphere. The results were obtained by comparing the light of the star and planet together with that of the star alone, when the planet moved behind the star.

Crime solvers start genetic portal

FORENSICS | Two leaders in using genetic genealogy to crack unsolved murders and other cold cases have launched a nonprofit to help law enforcement. Crime scene DNA doesn't always have a direct match in police databases, but genetic genealogists have

PHOTO: CLAUDIA MARCELLONI/CERN

tracked down murder and rape suspects and identified murder victims' remains by linking that DNA to relatives who have deposited their genetic profiles in online ancestry websites. Since the Golden State Killer was found this way in 2018, the technique has been used to solve hundreds of cases. But just two commercial genealogy databases allow law enforcement searches, and their fees are rising. To ensure low costs and stable access to people's genetic profiles, well-known genetic genealogists CeCe Moore and Margaret Press launched DNA Justice (dnajustice.org) on 12 March, inviting anyone to confidentially share their DNA data and family information. The database only had about 300 DNA profiles this week when *Science* went to press; it will need 100,000 or more for effective searches.

Marburg cases widen in Africa

EPIDEMIOLOGY | Public health experts are combating two outbreaks of the deadly Marburg virus on opposite sides of the African continent. Authorities in Equatorial Guinea have reported nine confirmed and 20 probable cases of the hemorrhagic fever since early January; 27 of the 29 patients have died. The cases are spread across different provinces, and health officials say the lack of connection between some cases suggests there is undetected spread in the community. In Tanzania, eight cases have been reported, including five deaths. This is the first Marburg outbreak in each country, and genome sequencing is ongoing to determine whether the two outbreaks are related. The risk of spread to other countries in the region is also high, the World Health Organization has warned. Unlike Ebola disease, to which it is related, Marburg has no approved vaccines or antivirals.

China's COVID-19 mRNA vaccine

INFECTIOUS DISEASES | China last week approved for emergency use its first COVID-19 vaccine using messenger RNA (mRNA) technology. The homegrown product, developed by CSPC Pharmaceutical Group, comes about 2 years after much of the rest of the world began to get shots based on the novel vaccine platform, developed in the United States and Europe. Trials with more than 5500 participants showed the vaccine is safe and effective, CSPC reported. China's Shanghai Fosun Pharmaceutical in March 2020 obtained the rights to market BioNTech's mRNA COVID-19 vaccine in China once it was developed, but regulators never approved it. China has instead relied on less effective, traditional vaccines made with inactivated coronaviruses.

U.S. autism prevalence rises

PUBLIC HEALTH | The prevalence of autism among U.S. 8-year-olds in 2020 reached its highest level since systematic surveillance began in 2000, the U.S. Centers for Disease Control and Prevention (CDC) reported last week. About 4% of boys and 1% of girls were affected—or one in 36 for all children, up from one in 44 in 2018, according to the analysis in the *Morbidity and Mortality Weekly Report* (*MMWR*). For the first time, autism's prevalence among 8-year-olds was higher in children from racial and ethnic minority groups than in white children, the analysis found, based on surveillance programs in 11 communities in 11 states. Autism's prevalence in 4-year-olds also grew, to one in 47 children in 2020, according to a separate analysis last week in *MMWR*. The COVID-19 pandemic disrupted autism diagnoses, these CDC data showed.



An illustration depicts ancient sea anemones about to be entombed by an underwater avalanche.

ILLUSTRATION: DENVER MUSEUM OF NATURE & SCIENCE

PALEONTOLOGY

Rare anemone fossils were mistaken as jellyfish

An abundant deposit of sea anemone fossils had been hidden in plain sight—until now. Well-preserved specimens of ancient soft-bodied organisms are rare in the fossil record, but researchers have identified a trove in northern Illinois's 310-million-year-old Mazon Creek deposit, they reported on 8 March in *Papers in Paleontology*. The fossils, known locally as "the blob," were misidentified in the 1970s as a species of jellyfish, formally named *Essexella asherae*. But the researchers realized their probable identity as sea anemones (the order *Actiniaria*) after reexamining thousands of museum specimens. Mazon Creek, renowned for its wealth of fossils, formed as an ancient river spewed sediment into a shallow sea. Underwater avalanches buried anemones and other soft-bodied animals, entombing them to become fossils.



IN DEPTH

ASTRONOMY

Earliest galaxies found by JWST confound theory

Too numerous and bright, they challenge ideas about star birth in the infant universe

By **Daniel Clery**

Charlotte Mason, an astrophysicist at the University of Copenhagen, had modest expectations 9 months ago, when she and her collaborators began to use JWST, the giant new space telescope, to look back in time for the universe's first galaxies. Modeling suggested the patch of sky they were examining would hold just 0.2 galaxies—none, in other words, unless they got lucky. Yet out of the images popped not one, but two bright galaxies. "That was the biggest surprise to me," she says.

The surprises have kept coming. JWST astronomers have found more than 15 galaxies shining within the first half-billion years of the 13.7-billion-year-old universe—far too many according to theorists' models of galaxy formation. Initial estimates of the galaxies' age and distance come from their brightness at particular wavelengths, but astronomers are now applying the gold standard method: analyzing the galaxies' spectra in detail to see how much their light has been stretched by the expansion of the universe. The spectra have confirmed nine of the early galaxies, including two added to the roster this week following JWST observations on 24 March. "This is the most exciting period of my recent life," Casey Papovich of Texas A&M University, College Station, told astronomers last week at a meeting at the Kavli Institute for Cosmology in Cambridge, England.

The discoveries are leaving theorists scratching their heads. The standard theory of cosmology, lambda-cold dark matter (ΛCDM), says clouds of dark matter—the mysterious stuff making up 85% of the universe's mass—began to clump up into halos soon after the big bang. The halos' strong gravity sucked in gas, which collapsed to form stars. ΛCDM cannot account for the excess galaxies astronomers are seeing, but few astronomers are ready to tear it up. "Let's get a bigger population," says Alice Shapley of the University of California, Los Angeles. "Then it will be time to look at theories."

Instead, cosmologists wonder whether the excess of galaxies in the newborn universe is more apparent than real. It could be that surveys so far have, by chance, zoomed in on areas dense with galaxies. The apparent excess could also arise if the galaxies are merely overly bright and stuffed with stars, so more of them poke above the threshold that JWST can see. But that creates a new theoretical problem: Why are they so bright and full of stars? "There's no convincing explanation yet," says Richard Ellis of University College London.

In young galaxies closer to Earth, feedbacks limit the rate of star formation. Theorists believe baby stars emit stellar winds: streams of particles that slow the process by blowing gas out of the galaxy. Adding to the effect are supernovae, which occur when fast-burning stars run out of fuel, collapsing and triggering explosions that blow away gas and surround the galaxy with dust, scattering its light and giving it a reddish hue. The galaxy's gravity draws some of the gas back in, but star formation efficiency, a measure of stars formed per unit of gas, typically sticks below 10%.

Avishai Dekel of the Hebrew University of Jerusalem argues that star formation must have been more efficient in the early universe, which was physically much smaller. The gas from which stars form would have been 1000 times denser than it is after billions of years of expansion, making star formation easier. Moreover, that primordial gas was not yet enriched with the heavier elements and dust forged by supernovae. As a result, the stellar winds of these first stars would have been less intense than today—and a weaker brake on star formation. For 1 million years or so, Dekel says, these galaxies could churn out stars with a formation efficiency of nearly 100%. "All galaxies

IMAGE: NASA; ESA; CSA; IVO LABBE/SWINBURNE; RACHEL BEZANSO

The JWST space telescope has zoomed in on a region called Pandora's Cluster in search of early galaxies.

at this epoch should make a feedback-free starburst if they are massive enough." What's more, the lack of dust would have allowed the stars to shine more brightly than comparable stars today, and at the bluer wavelengths seen by JWST.

Andrea Ferrara of the Scuola Normale Superiore in Pisa, Italy, takes a different tack. He says the dense galaxies of the early universe would ramp up star formation in cycles that repeat every 100 million years. During the star-forming phases, the radiation pressure from the stars would blast out dust, making the galaxies appear bright and blue.

Ferrara finds some evidence to support the model: The JWST spectrum of one distant galaxy, GNz-11, had one spectral line—for hydrogen gas—shifted out of place as if the gas was moving at 300 kilometers per second. "We see clear signs of outflowing material," with radiation pressure sweeping out both hydrogen and dust, he says. JWST has also spied a galaxy without signs of star formation 700 million years after the big bang, which Ferrara suggests could be in a quiet phase between star-forming bursts.

Another possible explanation for the galaxies' surprising brightness is that it was driven not by stars, but massive black holes at their hearts. Hot disks of dust and gas swirling down the gravitational drains of monster black holes are what drive quasars, some of the brightest objects in the universe. But astronomers have not seen quasars any earlier than about 650 million years after the big bang, and they struggle to explain how their black holes could have grown big enough to blaze brightly much earlier. Nonetheless, at the Kavli conference Papovich showed the JWST spectrum of a galaxy from when the universe was 550 million years old. It showed a hint of light being both stretched and squeezed—a telltale sign of swirling gases around a black hole. Ellis still isn't convinced giant black holes could form early enough. "The black hole idea is the most extreme," he says.

Few want to countenance an even more extreme option: that the ΛCDM model is at fault. It could be tweaked to produce more dark matter halos or larger ones able to concentrate gas more quickly into bigger galaxies. But theorists are loath to tinker with it because it explains so many things so well: the observed distribution of galaxies, the abundances of primordial gases, and the accelerating expansion of the universe. "We'd be at risk of screwing everything else up," Ferrara says. "You'd need to be pretty desperate." ■

PHOTO: LAUREN BISHOP/CDC

EVOLUTIONARY BIOLOGY

How the yellow fever mosquito found its first human victim

Aedes aegypti developed a taste for humans at the birth of the Sahara, a new analysis suggests

By **Joshua Sokol**

Some 500 years ago, a city-living, human-biting form of the yellow fever mosquito *Aedes aegypti* began to hitch rides out of West African ports during the Transatlantic Slave Trade. It spread to the Americas and then to Asia, causing centuries of disease outbreaks across the colonial world. Today, its globally invasive descendants act as the main disease vector for the yellow fever, Zika, chikungunya, and dengue viruses, together causing hundreds of millions of infections each year.

But how, exactly, the yellow fever mosquito first evolved to bite people, stow away on ships, and thrive in new destinations is murkier. Researchers agree on the story's outlines: A subpopulation of *A. aegypti* split from a harmless ancestor that preferred to live in forests and feed on animals, not people. "The thing we didn't know was when, how, or why all that happened," says Noah Rose, an evolutionary biologist at the University of California (UC), San Diego. Now, Rose has led a genomic analysis that concludes the fateful split occurred about 5000 years ago, during a period of natural climate change in the West African Sahel, at the Sahara's southern border.



Aedes aegypti may have developed its human-biting tastes 5000 years ago.

"We were able to really tell the natural story of *A. aegypti* using the genome," says Athanase Badolo, an entomologist at Joseph Ki-Zerbo University in Ouagadougou, Burkina Faso, who helped collect mosquitoes and co-authored the study. Those genomes also revealed a hint of rapid, ongoing evolution that in just the past few decades allowed the species to prey on humans and likely spread disease even more efficiently in African cities. About the findings, "The [researchers] have presented compelling evidence," says R. Nicolas Lou, a population geneticist at UC Berkeley who wasn't involved with the current study.

Rose's team began by sampling different yellow fever mosquito populations from forests and cities. In 2020, the researchers reported that mosquitoes with the strongest preferences for human odors seemed clustered in arid, urban communities in the Sahel, suggesting the first human-focused mosquitoes likely evolved there, drawn to towns offering dense human populations and water during long dry seasons.

For their newer analysis, Rose's team turned to a computational technique typically applied to reconstructing human migrations from divergent genomes scattered across the world. Once two biological populations are separated and can no longer interbreed, their genomes increasingly diverge over time. The accumulated mutations serve as a clock that, if calibrated with known dates, can be rewound to pinpoint the dates of divergences. In this case, graphs comparing mosquito genomes collected across Africa and in Brazil showed two intense migration events: one between Africa and the Americas, and a separate, much earlier one where all *A. aegypti* populations first split.

Ecologists believe the transatlantic migration of mosquitoes may have peaked around the year 1800, during the height of the Transatlantic Slave Trade. Along with some 80,000 enslaved people being stolen across the ocean each year, yellow fever mosquitoes

Check for updates

The JWST space telescope has zoomed in on a region called Pandora's Cluster in search of early galaxies.

at this epoch should make a feedback-free starburst if they are massive enough." What's more, the lack of dust would have allowed the stars to shine more brightly than comparable stars today, and at the bluer wavelengths seen by JWST.

Andrea Ferrara of the Scuola Normale Superiore in Pisa, Italy, takes a different tack. He says the dense galaxies of the early universe would ramp up star formation in cycles that repeat every 100 million years. During the star-forming phases, the radiation pressure from the stars would blast out dust, making the galaxies appear bright and blue.

Ferrara finds some evidence to support the model: The JWST spectrum of one distant galaxy, GNz-11, had one spectral line—for hydrogen gas—shifted out of place as if the gas was moving at 300 kilometers per second. "We see clear signs of outflowing material," with radiation pressure sweeping out both hydrogen and dust, he says. JWST has also spied a galaxy without signs of star formation 700 million years after the big bang, which Ferrara suggests could be in a quiet phase between star-forming bursts.

Another possible explanation for the galaxies' surprising brightness is that it was driven not by stars, but massive black holes at their hearts. Hot disks of dust and gas swirling down the gravitational drains of monster black holes are what drive quasars, some of the brightest objects in the universe. But astronomers have not seen quasars any earlier than about 650 million years after the big bang, and they struggle to explain how their black holes could have grown big enough to blaze brightly much earlier. Nonetheless, at the Kavli conference Papovich showed the JWST spectrum of a galaxy from when the universe was 550 million years old. It showed a hint of light being both stretched and squeezed—a telltale sign of swirling gases around a black hole. Ellis still isn't convinced giant black holes could form early enough. "The black hole idea is the most extreme," he says.

Few want to countenance an even more extreme option: that the ΛCDM model is at fault. It could be tweaked to produce more dark matter halos or larger ones able to concentrate gas more quickly into bigger galaxies. But theorists are loath to tinker with it because it explains so many things so well: the observed distribution of galaxies, the abundances of primordial gases, and the accelerating expansion of the universe. "We'd be at risk of screwing everything else up," Ferrara says. "You'd need to be pretty desperate." ■

PHOTO: LAUREN BISHOP/CDC

EVOLUTIONARY BIOLOGY

How the yellow fever mosquito found its first human victim

Aedes aegypti developed a taste for humans at the birth of the Sahara, a new analysis suggests

By **Joshua Sokol**

Some 500 years ago, a city-living, human-biting form of the yellow fever mosquito *Aedes aegypti* began to hitch rides out of West African ports during the Transatlantic Slave Trade. It spread to the Americas and then to Asia, causing centuries of disease outbreaks across the colonial world. Today, its globally invasive descendants act as the main disease vector for the yellow fever, Zika, chikungunya, and dengue viruses, together causing hundreds of millions of infections each year.

But how, exactly, the yellow fever mosquito first evolved to bite people, stow away on ships, and thrive in new destinations is murkier. Researchers agree on the story's outlines: A subpopulation of *A. aegypti* split from a harmless ancestor that preferred to live in forests and feed on animals, not people. "The thing we didn't know was when, how, or why all that happened," says Noah Rose, an evolutionary biologist at the University of California (UC), San Diego. Now, Rose has led a genomic analysis that concludes the fateful split occurred about 5000 years ago, during a period of natural climate change in the West African Sahel, at the Sahara's southern border.



Aedes aegypti may have developed its human-biting tastes 5000 years ago.

"We were able to really tell the natural story of *A. aegypti* using the genome," says Athanase Badolo, an entomologist at Joseph Ki-Zerbo University in Ouagadougou, Burkina Faso, who helped collect mosquitoes and co-authored the study. Those genomes also revealed a hint of rapid, ongoing evolution that in just the past few decades allowed the species to prey on humans and likely spread disease even more efficiently in African cities. About the findings, "The [researchers] have presented compelling evidence," says R. Nicolas Lou, a population geneticist at UC Berkeley who wasn't involved with the current study.

Rose's team began by sampling different yellow fever mosquito populations from forests and cities. In 2020, the researchers reported that mosquitoes with the strongest preferences for human odors seemed clustered in arid, urban communities in the Sahel, suggesting the first human-focused mosquitoes likely evolved there, drawn to towns offering dense human populations and water during long dry seasons.

For their newer analysis, Rose's team turned to a computational technique typically applied to reconstructing human migrations from divergent genomes scattered across the world. Once two biological populations are separated and can no longer interbreed, their genomes increasingly diverge over time. The accumulated mutations serve as a clock that, if calibrated with known dates, can be rewound to pinpoint the dates of divergences. In this case, graphs comparing mosquito genomes collected across Africa and in Brazil showed two intense migration events: one between Africa and the Americas, and a separate, much earlier one where all *A. aegypti* populations first split.

Ecologists believe the transatlantic migration of mosquitoes may have peaked around the year 1800, during the height of the Transatlantic Slave Trade. Along with some 80,000 enslaved people being stolen across the ocean each year, yellow fever mosquitoes

stowed away on these voyages, laying their eggs in water barrels and feasting on the people aboard.

That timing helped Rose calibrate the date of the earlier migration event, when all *A. aegypti* insects first split into the forest- and city-dwelling types. "That's just clever," says Sadie Ryan, a medical geographer at the University of Florida who wasn't involved in the study, now accepted in the journal *eLife*.

That 5000-year figure for the emergence of human-loving yellow fever mosquitoes fits into a compelling ecological picture, Rose says. Right at that time, a drying climate was transforming the Sahara from grasslands to desert. As water grew scarce, a subset of mosquitoes may have adapted to lay their eggs in water storage containers in communities along the desert's edge. The mosquitoes then switched from feeding opportunistically on any animals around to the most plentiful food sources in this new habitat: humans.

A. aegypti continues to evolve. Rose's time-rewinding genomic analysis also revealed one more, ongoing evolutionary event in rapidly urbanizing environments like Ouagadougou. It's a region dominated by ancestral, animal-biting mosquitoes. But in the past 20 to 40 years, over the same period that many African cities saw explosive growth, human blood–seeking mosquitoes have interbred with more benign ones. Now, genes associated with the human-adapted mosquitoes are surging, likely because they confer an advantage in an ever-changing habitat.

The evolving mosquitoes might explain disease transmission patterns, Rose says. For instance, Burkina Faso had its first modern dengue outbreak in 2016, and the disease has returned every year since. And Ouagadougou's mosquitoes are changing still: Entomologists behind a forthcoming study found *A. aegypti* had quickly adapted to breed in the public hand-washing stations installed during the COVID-19 pandemic. "I think the situation is going to be worse in the coming years," Badolo says.

Studies like this one, using science and history reciprocally, are a welcome development, says J. R. McNeill, an environmental historian at Georgetown University who has argued that yellow fever mosquitoes reshaped Caribbean colonial history for 2 centuries. Yellow fever, he said, gave local armies—like the one commanded by Toussaint L'Ouverture to fight for Haiti's liberation—a crucial edge over invaders with immune systems unused to the virus. "I find it a really exciting turn in the historical profession and more broadly, the entire community of scholars and scientists who are interested in the past." ■



For centuries local people made pilgrimages to the former Maya city of Yaxchilán.

PHOTO: JON G. FULLER/VWPICS VIA AP IMAGES

ARCHAEOLOGY

Archaeologists study what ruins meant to people in the past

Mesoamerican sites offer insights into how ancient people viewed their own history

By **Lizzie Wade**

Around 500 C.E., a new government arose in the community now called Río Viejo, near the coast of the Mexican state of Oaxaca. It was once the largest city in the region, but it had shrunk by half and lost its political authority. The new rulers aimed to step into that power vacuum. But they had one problem: the ruins of a complex of ceremonial buildings built by Río Viejo's last centralized government centuries earlier. When that government collapsed, the temples and plazas had been ritually burned and left to decay, a reminder that hierarchical rulership had already failed once in Río Viejo. How would the new leaders manage the threat it posed?

Arthur Joyce, an archaeologist at the University of Colorado (CU), Boulder, has found they did so by putting their stamp on the ruins with a massive offering and portraits of themselves, set on top of the eroded surface of the old buildings. "These new rulers may have been trying to assert control over this thing that by its very existence would have questioned the inevitability and legitimacy of their power," Joyce says.

Previous generations of researchers tended to treat the massive ruins that dot Mexico and Central America as "inconsequential" in the lives of the people who lived nearby in later periods, Joyce says. Once a site emptied out and started to crumble, archaeologists typically concluded its importance had faded for people in the past. But a growing number are now recognizing that for people in precolonial Mesoamerica, "ruins, ancient objects, and ancestors were active parts of their communities," says Roberto Rosado-Ramirez, an archaeologist at Northwestern University.

In a session he and Joyce are organizing at this week's conference of the Society for American Archaeology (SAA) in Portland, Oregon, researchers will share new findings and ideas about ruins' roles in ancient Mesoamerican communities. "People in the past had their own past," says Christina Halperin, an archaeologist at the University of Montreal. By looking at how people interacted with the ruins around them, archaeologists can get a glimpse of how those communities conceived of their own history.

In Europe, archaeologists and historians have long studied the role of ancient

stowed away on these voyages, laying their eggs in water barrels and feasting on the people aboard.

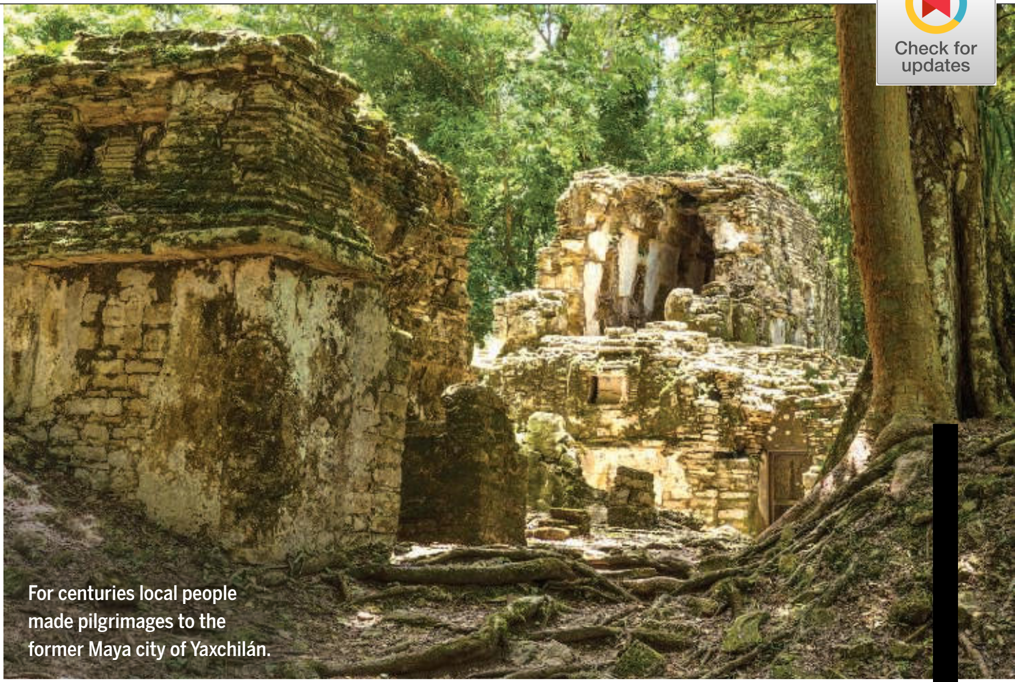
That timing helped Rose calibrate the date of the earlier migration event, when all *A. aegypti* insects first split into the forest- and city-dwelling types. "That's just clever," says Sadie Ryan, a medical geographer at the University of Florida who wasn't involved in the study, now accepted in the journal *eLife*.

That 5000-year figure for the emergence of human-loving yellow fever mosquitoes fits into a compelling ecological picture, Rose says. Right at that time, a drying climate was transforming the Sahara from grasslands to desert. As water grew scarce, a subset of mosquitoes may have adapted to lay their eggs in water storage containers in communities along the desert's edge. The mosquitoes then switched from feeding opportunistically on any animals around to the most plentiful food sources in this new habitat: humans.

A. aegypti continues to evolve. Rose's time-rewinding genomic analysis also revealed one more, ongoing evolutionary event in rapidly urbanizing environments like Ouagadougou. It's a region dominated by ancestral, animal-biting mosquitoes. But in the past 20 to 40 years, over the same period that many African cities saw explosive growth, human blood–seeking mosquitoes have interbred with more benign ones. Now, genes associated with the human-adapted mosquitoes are surging, likely because they confer an advantage in an ever-changing habitat.

The evolving mosquitoes might explain disease transmission patterns, Rose says. For instance, Burkina Faso had its first modern dengue outbreak in 2016, and the disease has returned every year since. And Ouagadougou's mosquitoes are changing still: Entomologists behind a forthcoming study found *A. aegypti* had quickly adapted to breed in the public hand-washing stations installed during the COVID-19 pandemic. "I think the situation is going to be worse in the coming years," Badolo says.

Studies like this one, using science and history reciprocally, are a welcome development, says J. R. McNeill, an environmental historian at Georgetown University who has argued that yellow fever mosquitoes reshaped Caribbean colonial history for 2 centuries. Yellow fever, he said, gave local armies—like the one commanded by Toussaint L'Ouverture to fight for Haiti's liberation—a crucial edge over invaders with immune systems unused to the virus. "I find it a really exciting turn in the historical profession and more broadly, the entire community of scholars and scientists who are interested in the past." ■



For centuries local people made pilgrimages to the former Maya city of Yaxchilán.

ARCHAEOLOGY

Archaeologists study what ruins meant to people in the past

Mesoamerican sites offer insights into how ancient people viewed their own history

By **Lizzie Wade**

Around 500 C.E., a new government arose in the community now called Río Viejo, near the coast of the Mexican state of Oaxaca. It was once the largest city in the region, but it had shrunk by half and lost its political authority. The new rulers aimed to step into that power vacuum. But they had one problem: the ruins of a complex of ceremonial buildings built by Río Viejo's last centralized government centuries earlier. When that government collapsed, the temples and plazas had been ritually burned and left to decay, a reminder that hierarchical rulership had already failed once in Río Viejo. How would the new leaders manage the threat it posed?

Arthur Joyce, an archaeologist at the University of Colorado (CU), Boulder, has found they did so by putting their stamp on the ruins with a massive offering and portraits of themselves, set on top of the eroded surface of the old buildings. "These new rulers may have been trying to assert control over this thing that by its very existence would have questioned the inevitability and legitimacy of their power," Joyce says.

Previous generations of researchers tended to treat the massive ruins that dot Mexico and Central America as "inconsequential" in the lives of the people who lived nearby in later periods, Joyce says. Once a site emptied out and started to crumble, archaeologists typically concluded its importance had faded for people in the past. But a growing number are now recognizing that for people in precolonial Mesoamerica, "ruins, ancient objects, and ancestors were active parts of their communities," says Roberto Rosado-Ramirez, an archaeologist at Northwestern University.

In a session he and Joyce are organizing at this week's conference of the Society for American Archaeology (SAA) in Portland, Oregon, researchers will share new findings and ideas about ruins' roles in ancient Mesoamerican communities. "People in the past had their own past," says Christina Halperin, an archaeologist at the University of Montreal. By looking at how people interacted with the ruins around them, archaeologists can get a glimpse of how those communities conceived of their own history.

In Europe, archaeologists and historians have long studied the role of ancient

PHOTO: JON G. FULLER/VWPICS VIA AP IMAGES

monuments such as Stonehenge in the cultures of later people. But in Mesoamerica and other colonized places, European settlers—and archaeologists—"pretended that the people they were colonizing had no real history, and therefore no claims to their land," says Shannon Dawdy, an anthropologist at the University of Chicago and a discussant in the SAA session. Studying ruins in the past is a way to center Indigenous perspectives about history that researchers previously ignored or denied, Rosado-Ramirez says.

Those ruins possessed spiritual, historical, and political power, which ambitious elites sometimes tried to manipulate for their own ends. At Río Viejo, Joyce found the rulers who took control around 500 C.E. placed their elaborate offering right in the center of the old complex, now called the Acropolis. The offering consisted of several burials, including three bodies interred within large ceramic vessels, and burned human and animal bones that indicate sacrifice, all topped with a layer of earth and two stone monuments called stelae. Three carved stone monuments dedicated to the new rulers, complete with their names and, in two cases, their portraits, were installed just to the north. "They attempted to appropriate [the Acropolis] by … putting their images on it," Joyce says.

> **"People in the past had their own past."**
> **Christina Halperin**, University of Montreal

In Ucanal, a Maya city in Guatemala where Halperin works, leaders may have used ruins to mark a break with the past, rather than try to reclaim it. In the early ninth century C.E., when many nearby cities were collapsing, writing on monuments shows a new regime took over in Ucanal. It tore down many older buildings and used their stones to construct new ones, but a few, including a small ballcourt, were left as ruins in the middle of the city. "It seems like they are purposely trying to generate a new era of their own history," Halperin says. It worked. Ucanal not only survived a time of widespread political crisis, but grew bigger than ever.

In other places, commoners, not elites, engaged with ruins. Around the time Ucanal was thriving under new leadership, the Maya city of El Perú-Waka' was faltering. The royal family had lost power, many people were moving away, and the remaining residents were struggling to maintain the city's centuries-old main temple. Olivia Navarro-Farr, an archaeologist at the College of Wooster, found offerings of everyday objects, such as domestic pottery and stone tools, deposited on either side of the temple staircase during the time the city was being abandoned.

When the city was at its height, the temple was home to royal tombs and stone monuments dedicated to past rulers. These late offerings are "the same materials [common] people used in their own homes," Navarro-Farr says. She thinks they were a way for common people to honor the building's long history and their ancestral connection to it during a time of political and social upheaval that may have threatened those bonds.

Many Indigenous communities in Mexico and Central America continue to leave offerings to ancestors, forest spirits, and deities at ancient sites, but archaeologists haven't always respected the tradition. Josuhé Lozada Toledo, an archaeologist at Mexico's National Institute of Anthropology and History, works with Lacandon Maya communities in the Mexican state of Chiapas. For centuries, they made pilgrimages to the ancient Maya city of Yaxchilán near the Mexico-Guatemala border, which they consider the home of the creator god Hachakyum. But they stopped in the 1980s, when more archaeologists and tourists began to visit Yaxchilán and practices such as burning incense inside the temples were prohibited, Lozada Toledo says. Few Lacandon people now remember the pilgrimage routes. At the SAA session, Lozada Toledo's team will describe how it is trying to help preserve them using ethnography and mapping software.

For contemporary Maya people, the remains of ancient buildings and settlements are never empty, even though they may no longer be occupied by humans, says Genner Llanes-Ortiz, a social anthropologist at Bishop's University who is Yucatec Maya. "They are sacred places, they are inhabited places, and they are places with their own character," he says.

Because of this, Llanes-Ortiz, who is not participating in the SAA session, questions the use of the term "ruin" to describe older sites. "I appreciate [the session's] approach of trying to interrogate the Eurocentric idea that 'ruins' are just decayed buildings," he says. But words for these sites should emphasize their active, vibrant role in present and past communities. In Yucatec Maya, for example, an archaeological site is often called a *múul*, or hill. "Mountains or hills are also agents in the life of the community," with a personhood similar to human beings, Llanes-Ortiz says.

"A lot of us [archaeologists] tend to think that Western ways of seeing the world are somehow natural or obvious, or the only option," says Sarah Kurnick, an archaeologist at CU who is participating in the session. "I'm looking forward to seeing an elevation of non-Western viewpoints as critical to understanding the past." ■

PUBLISHING

Fast-growing open-access journals lose impact factors

Web of Science delists some 50 journals, including one of the world's largest

By **Jeffrey Brainard**

Nearly two dozen journals from two of the fastest growing open-access publishers, including one of the world's largest journals by volume, will no longer receive a key scholarly imprimatur. On 20 March, the Web of Science database said it delisted the journals along with dozens of others, stripping them of an impact factor, the citation-based measure of quality that, although controversial, carries weight with authors and institutions. The move highlights continuing debate about a business model marked by high volumes of articles, ostensibly chosen for scientific soundness rather than novelty, and the practice by some open-access publishers of recruiting large numbers of articles for guest-edited special issues.

The Web of Science Master Journal List, run by the analytics company Clarivate, lists journals based on 24 measures of quality, including effective peer review and adherence to ethical publishing practices, and periodically checks that listed journals meet the standards. Clarivate calculates impact factors for a select subset of journals on the list. The company expanded quality checks this year because of "increasing threats to the integrity of the scholarly record," Web of Science's Editor-in-Chief Nandita Quaderi says. The company removed 50 journals from the list, an unusually large number for a single year, and Clarivate said it is continuing to review 450 more.

"My expectation is that this initial delisting … is only the tip of the iceberg," says Rob Johnson, managing director of Research Consulting, a firm that advises science publishers and funders.

Clarivate initially did not name any of the delisted journals or provide specific reasons. But it confirmed to *Science* the identities of 19 Hindawi journals and two MDPI

monuments such as Stonehenge in the cultures of later people. But in Mesoamerica and other colonized places, European settlers—and archaeologists—"pretended that the people they were colonizing had no real history, and therefore no claims to their land," says Shannon Dawdy, an anthropologist at the University of Chicago and a discussant in the SAA session. Studying ruins in the past is a way to center Indigenous perspectives about history that researchers previously ignored or denied, Rosado-Ramirez says.

Those ruins possessed spiritual, historical, and political power, which ambitious elites sometimes tried to manipulate for their own ends. At Río Viejo, Joyce found the rulers who took control around 500 C.E. placed their elaborate offering right in the center of the old complex, now called the Acropolis. The offering consisted of several burials, including three bodies interred within large ceramic vessels, and burned human and animal bones that indicate sacrifice, all topped with a layer of earth and two stone monuments called stelae. Three carved stone monuments dedicated to the new rulers, complete with their names and, in two cases, their portraits, were installed just to the north. "They attempted to appropriate [the Acropolis] by … putting their images on it," Joyce says.

In Ucanal, a Maya city in Guatemala where Halperin works, leaders may have used ruins to mark a break with the past, rather than try to reclaim it. In the early ninth century C.E., when many nearby cities were collapsing, writing on monuments shows a new regime took over in Ucanal. It tore down many older buildings and used their stones to construct new ones, but a few, including a small ballcourt, were left as ruins in the middle of the city. "It seems like they are purposely trying to generate a new era of their own history," Halperin says. It worked. Ucanal not only survived a time of widespread political crisis, but grew bigger than ever.

In other places, commoners, not elites, engaged with ruins. Around the time Ucanal was thriving under new leadership, the Maya city of El Perú-Waka' was faltering. The royal family had lost power, many people were moving away, and the remaining residents were struggling to maintain the city's centuries-old main temple. Olivia Navarro-Farr, an archaeologist at the College of Wooster, found offerings of everyday objects, such as domestic pottery and stone tools, deposited on either side of the temple staircase during the time the city was being abandoned.

When the city was at its height, the temple was home to royal tombs and stone monuments dedicated to past rulers. These late offerings are "the same materials [common] people used in their own homes," Navarro-Farr says. She thinks they were a way for common people to honor the building's long history and their ancestral connection to it during a time of political and social upheaval that may have threatened those bonds.

> **"People in the past had their own past."**
> **Christina Halperin**, University of Montreal

Many Indigenous communities in Mexico and Central America continue to leave offerings to ancestors, forest spirits, and deities at ancient sites, but archaeologists haven't always respected the tradition. Josuhé Lozada Toledo, an archaeologist at Mexico's National Institute of Anthropology and History, works with Lacandon Maya communities in the Mexican state of Chiapas. For centuries, they made pilgrimages to the ancient Maya city of Yaxchilán near the Mexico-Guatemala border, which they consider the home of the creator god Hachakyum. But they stopped in the 1980s, when more archaeologists and tourists began to visit Yaxchilán and practices such as burning incense inside the temples were prohibited, Lozada Toledo says. Few Lacandon people now remember the pilgrimage routes. At the SAA session, Lozada Toledo's team will describe how it is trying to help preserve them using ethnography and mapping software.

For contemporary Maya people, the remains of ancient buildings and settlements are never empty, even though they may no longer be occupied by humans, says Genner Llanes-Ortiz, a social anthropologist at Bishop's University who is Yucatec Maya. "They are sacred places, they are inhabited places, and they are places with their own character," he says.

Because of this, Llanes-Ortiz, who is not participating in the SAA session, questions the use of the term "ruin" to describe older sites. "I appreciate [the session's] approach of trying to interrogate the Eurocentric idea that 'ruins' are just decayed buildings," he says. But words for these sites should emphasize their active, vibrant role in present and past communities. In Yucatec Maya, for example, an archaeological site is often called a *múul*, or hill. "Mountains or hills are also agents in the life of the community," with a personhood similar to human beings, Llanes-Ortiz says.

"A lot of us [archaeologists] tend to think that Western ways of seeing the world are somehow natural or obvious, or the only option," says Sarah Kurnick, an archaeologist at CU who is participating in the session. "I'm looking forward to seeing an elevation of non-Western viewpoints as critical to understanding the past." ■

PUBLISHING

Fast-growing open-access journals lose impact factors

Web of Science delists some 50 journals, including one of the world's largest

By **Jeffrey Brainard**

Nearly two dozen journals from two of the fastest growing open-access publishers, including one of the world's largest journals by volume, will no longer receive a key scholarly imprimatur. On 20 March, the Web of Science database said it delisted the journals along with dozens of others, stripping them of an impact factor, the citation-based measure of quality that, although controversial, carries weight with authors and institutions. The move highlights continuing debate about a business model marked by high volumes of articles, ostensibly chosen for scientific soundness rather than novelty, and the practice by some open-access publishers of recruiting large numbers of articles for guest-edited special issues.

The Web of Science Master Journal List, run by the analytics company Clarivate, lists journals based on 24 measures of quality, including effective peer review and adherence to ethical publishing practices, and periodically checks that listed journals meet the standards. Clarivate calculates impact factors for a select subset of journals on the list. The company expanded quality checks this year because of "increasing threats to the integrity of the scholarly record," Web of Science's Editor-in-Chief Nandita Quaderi says. The company removed 50 journals from the list, an unusually large number for a single year, and Clarivate said it is continuing to review 450 more.

"My expectation is that this initial delisting … is only the tip of the iceberg," says Rob Johnson, managing director of Research Consulting, a firm that advises science publishers and funders.

Clarivate initially did not name any of the delisted journals or provide specific reasons. But it confirmed to *Science* the identities of 19 Hindawi journals and two MDPI

titles after reports circulated about their removals. The MDPI journals include the *International Journal of Environmental Research and Public Health*, which published about 17,000 articles last year.

Clarivate told *Science* it removed the MDPI journals based on Web of Science's "content relevance" criterion. MDPI's practice of publishing large numbers of special issues is likely at the heart of the concern, outsiders say. "[Clarivate's] announcement suggests we are approaching the high-water mark for the use of special issues as a growth model," Johnson says.

Articles in special issues account for most of the meteoric growth of MDPI since it was founded in 2010. Formerly called the Multidisciplinary Digital Publishing Institute, it has since become the largest publisher of open-access papers and the fourth-largest publisher of scholarly papers overall, producing some 400 journals. In 2022, nearly 100 of those journals that have impact factors published more than 17,000 special issues, containing 187,000 articles, according to an unpublished analysis by Paolo Crosetto, an economist at France's National Research Institute for Agriculture, Food, and the Environment, and Pablo Gómez Barreiro of the Royal Botanic Gardens, Kew. Some MDPI titles published four special issues a day.

Skeptics worry the practice is especially vulnerable to manipulation by guest editors who lack expertise, have conflicts of interest, or accept fabricated manuscripts produced by paper mills. "I have no proof that they did anything wrong," Crosetto says, adding that guest editor practices appear to vary. "But it stands to reason that trust [in the product's quality] is hard when you leave this guest editing to anyone."

Carlos Peixeira Marques of the University of Trás-os-Montes and Alto Douro, for example, says MDPI sent him multiple invitations to serve as a guest editor, in agriculture, animal science, and engineering—but never in his field of business and tourism. "The absolutely insane number of [MDPI] special issues has made it impossible to guarantee minimum peer-review standards," he says.

The speed with which MDPI's special issue manuscripts are reviewed and published is also a concern, Crosetto says. In 2022, MDPI's median time from submission to acceptance was 37 days, well below the 200 days at the PLOS family of journals, another large, open-access publisher he examined for comparison. For about one in three papers, MDPI's turnaround was 1 month or less. Considering the time it can take to recruit reviewers and revise manuscripts, "this looks just impossible," he says.

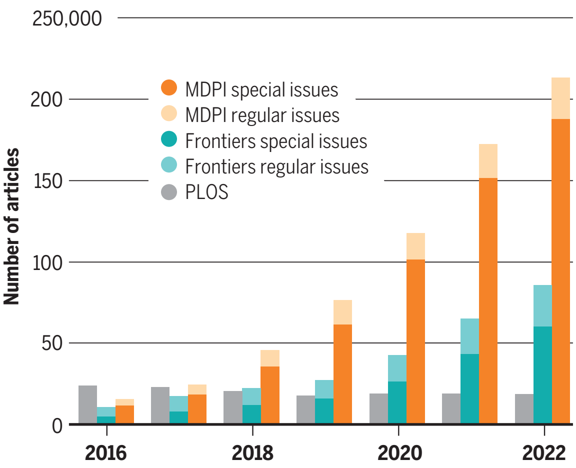
Attempts by *Science* to reach MDPI for comment were unsuccessful. But in past statements, the company and supporters have said fast peer review, done properly, can allow authors to share results with colleagues more quickly than via traditional journals. Serving as a guest editor can help junior researchers develop editing skills and network with colleagues. And the company says it provides an outlet for technically sound papers that might be rejected by more selective journals.

Special issues have been also been problematic for Hindawi. Its owner, scholarly publisher Wiley, announced on 9 March that it suspended publishing special issues in Hindawi journals from mid-October 2022 to mid-January after it identified "compromised articles." The special issues were targeted by paper mills and "bad actors" who fabricated content, says Jay Flynn, Wiley's executive vice president.

In response, Wiley added staffing and increased editorial controls, pairing artificial intelligence–based screening with manual checks, he said. This led to 500 retractions in Hindawi journals, with an estimated 1200 more to come. Hindawi reported a loss of $9 million in revenue from the pause in special issues for the quarter ending in January. The 19 Hindawi journals delisted by Clarivate represent about one-third of its journals that had been listed in Web of Science. ■

Ever so special

Articles in special issues with guest editors have driven much of the growth of open-access publisher MDPI. Numbers for two other open-access publishers are shown for comparison.



Data for Hindawi were not included in this analysis.

ARCHAEOLOGY

'Persian princes' helped found early African trade power

Ancient DNA reflects foreign influx to Swahili coast, but its culture is largely local

By **Andrew Curry**

The Swahili coast, stretching more than 3000 kilometers from southern Ethiopia to Tanzania, was a hub of medieval trade, exporting ivory and other resources from the African interior to South Asia, the Arab world, and Persia. Its cultural legacy remains potent: Swahili is now spoken across large parts of Africa, and the ruins of ancient towns, many with mosques and other buildings cut from shoreline coral deposits, record the coast's heyday. But whether Swahili culture was indigenous to Africa or arrived from overseas has been an ongoing debate.

One seemingly fanciful account dates from the 1500s, when Arab chroniclers recorded the stories Swahili people told about their origins. According to one version, known as the Kilwa Chronicle, seven Persian princes fleeing persecution set sail from the trading hub of Shiraz. After washing up on the coast of Africa, they founded a dynasty that ruled the Swahili coast for centuries.

An analysis of 54 genomes from people buried in Swahili coastal towns between 1250 and 1800 C.E. now gives that tale scientific support—while showing much of Swahili culture was derived from local African ancestors. The DNA of medieval people buried in elite Swahili cemeteries around 1200 C.E. shows their male forebears were closely related to people in modern-day Iran. Their female ancestors, meanwhile, were almost entirely local, with genomes resembling Bantu groups living in the region today.

University of South Florida archaeologist Chapurukha Kusimba, who led the study, published this week in *Nature*, believes it finally resolves the mysterious history of the Swahili coast. "This long-standing question has been answered," he says.

Gathered at seven sites in modern-day Kenya and Tanzania, the data represent the

CREDITS: (GRAPHIC) C. BICKEL/*SCIENCE*; (DATA) PAOLO CROSETTO/

titles after reports circulated about their removals. The MDPI journals include the *International Journal of Environmental Research and Public Health*, which published about 17,000 articles last year.

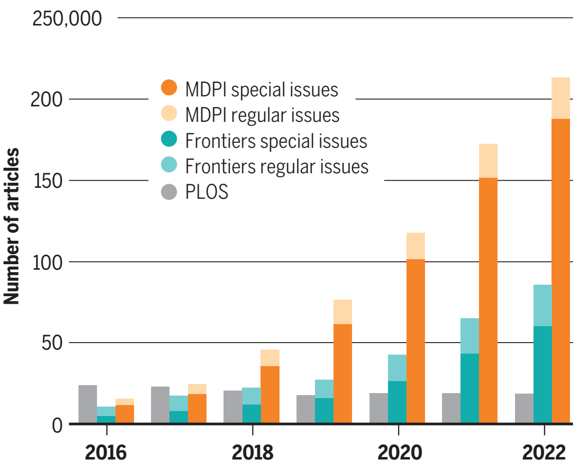
Clarivate told *Science* it removed the MDPI journals based on Web of Science's "content relevance" criterion. MDPI's practice of publishing large numbers of special issues is likely at the heart of the concern, outsiders say. "[Clarivate's] announcement suggests we are approaching the high-water mark for the use of special issues as a growth model," Johnson says.

Articles in special issues account for most of the meteoric growth of MDPI since it was founded in 2010. Formerly called the Multidisciplinary Digital Publishing Institute, it has since become the largest publisher of open-access papers and the fourth-largest publisher of scholarly papers overall, producing some 400 journals. In 2022, nearly 100 of those journals that have impact factors published more than 17,000 special issues, containing 187,000 articles, according to an unpublished analysis by Paolo Crosetto, an economist at France's National Research Institute for Agriculture, Food, and the Environment, and Pablo Gómez Barreiro of the Royal Botanic Gardens, Kew. Some MDPI titles published four special issues a day.

Skeptics worry the practice is especially vulnerable to manipulation by guest editors who lack expertise, have conflicts of interest, or accept fabricated manuscripts produced by paper mills. "I have no proof that they did anything wrong," Crosetto says, adding that guest editor practices appear to vary. "But it stands to reason that trust [in the product's quality] is hard when you leave this guest editing to anyone."

Ever so special

Articles in special issues with guest editors have driven much of the growth of open-access publisher MDPI. Numbers for two other open-access publishers are shown for comparison.



Data for Hindawi were not included in this analysis.

Carlos Peixeira Marques of the University of Trás-os-Montes and Alto Douro, for example, says MDPI sent him multiple invitations to serve as a guest editor, in agriculture, animal science, and engineering—but never in his field of business and tourism. "The absolutely insane number of [MDPI] special issues has made it impossible to guarantee minimum peer-review standards," he says.

The speed with which MDPI's special issue manuscripts are reviewed and published is also a concern, Crosetto says. In 2022, MDPI's median time from submission to acceptance was 37 days, well below the 200 days at the PLOS family of journals, another large, open-access publisher he examined for comparison. For about one in three papers, MDPI's turnaround was 1 month or less. Considering the time it can take to recruit reviewers and revise manuscripts, "this looks just impossible," he says.

Attempts by *Science* to reach MDPI for comment were unsuccessful. But in past statements, the company and supporters have said fast peer review, done properly, can allow authors to share results with colleagues more quickly than via traditional journals. Serving as a guest editor can help junior researchers develop editing skills and network with colleagues. And the company says it provides an outlet for technically sound papers that might be rejected by more selective journals.

Special issues have been also been problematic for Hindawi. Its owner, scholarly publisher Wiley, announced on 9 March that it suspended publishing special issues in Hindawi journals from mid-October 2022 to mid-January after it identified "compromised articles." The special issues were targeted by paper mills and "bad actors" who fabricated content, says Jay Flynn, Wiley's executive vice president.

In response, Wiley added staffing and increased editorial controls, pairing artificial intelligence–based screening with manual checks, he said. This led to 500 retractions in Hindawi journals, with an estimated 1200 more to come. Hindawi reported a loss of $9 million in revenue from the pause in special issues for the quarter ending in January. The 19 Hindawi journals delisted by Clarivate represent about one-third of its journals that had been listed in Web of Science. ■

ARCHAEOLOGY

'Persian princes' helped found early African trade power

Ancient DNA reflects foreign influx to Swahili coast, but its culture is largely local

By **Andrew Curry**

The Swahili coast, stretching more than 3000 kilometers from southern Ethiopia to Tanzania, was a hub of medieval trade, exporting ivory and other resources from the African interior to South Asia, the Arab world, and Persia. Its cultural legacy remains potent: Swahili is now spoken across large parts of Africa, and the ruins of ancient towns, many with mosques and other buildings cut from shoreline coral deposits, record the coast's heyday. But whether Swahili culture was indigenous to Africa or arrived from overseas has been an ongoing debate.

One seemingly fanciful account dates from the 1500s, when Arab chroniclers recorded the stories Swahili people told about their origins. According to one version, known as the Kilwa Chronicle, seven Persian princes fleeing persecution set sail from the trading hub of Shiraz. After washing up on the coast of Africa, they founded a dynasty that ruled the Swahili coast for centuries.

An analysis of 54 genomes from people buried in Swahili coastal towns between 1250 and 1800 C.E. now gives that tale scientific support—while showing much of Swahili culture was derived from local African ancestors. The DNA of medieval people buried in elite Swahili cemeteries around 1200 C.E. shows their male forebears were closely related to people in modern-day Iran. Their female ancestors, meanwhile, were almost entirely local, with genomes resembling Bantu groups living in the region today.

University of South Florida archaeologist Chapurukha Kusimba, who led the study, published this week in *Nature*, believes it finally resolves the mysterious history of the Swahili coast. "This long-standing question has been answered," he says.

Gathered at seven sites in modern-day Kenya and Tanzania, the data represent the

CREDITS: (GRAPHIC) C. BICKEL/*SCIENCE*; (DATA) PAOLO CROSETTO/

largest ancient DNA study yet from an African context. Combined with archaeological evidence from towns all along the Swahili coast and genetic evidence from people living there today, "It's really an extraordinary piece of scholarship," says Peter Schmidt, an archaeologist at the University of Florida who was not involved in the research.

The study does not support the simple picture that colonial-era British archaeologists favored. "The dominant paradigm was that this was a foreign civilization, with African involvement," Schmidt says. "The idea was Persians or Arabs brought civilization with them to benighted, primitive Africans," adds Mark Horton, an archaeologist at the Royal Agricultural University.

A postcolonial backlash posited the opposite, arguing that the medieval Swahili culture was entirely African in origin. The architecture of the "stone towns" was distinct from foreign styles, and Swahili was clearly a Bantu language, with loan words from Arabic, Persian, Portuguese, and other languages from overseas.

When Kusimba started to excavate the cemeteries of Swahili towns in the mid-1990s, what he found supported that picture. Of the artifacts he recovered, 95% were of local origin, he says, with only a few imported trade goods. But Kusimba decided to search for more direct evidence about the origins of the Swahili founders. "The people who lived and died in these towns were still there—why don't we dig them out and examine them?" Kusimba, who is originally from Kenya, worked with local communities to excavate the human remains, analyze them, and rebury them.

The skeletons were similar in build to people buried farther inland, a clue that they were local in origin. Kusimba thought DNA extracted from the bones might tell a clearer story. But 20 years ago ancient DNA techniques were in their infancy; many researchers thought warm regions such as eastern Africa would never yield useful data because hot weather degrades genetic material. "The only way to answer these questions was to do archaeogenetics, but its time had not come," Kusimba says.

Improvements in DNA sampling techniques and more powerful analytical capabilities changed that. In the new study, DNA from medieval cemeteries used by the Swahili elite revealed a genetic influx from Persia that was "overwhelmingly male," according to co-author Esther Brielle, a geneticist at Harvard University. Some of the individuals derived more than 70% of their male-line ancestry from outside Africa, in contrast to their African female ancestors. "The sex bias was a surprise," she says. "With results like that, we often think there must have been male marauders coming in for conquest."

To collaborators familiar with Swahili culture past and present, however, a violent takeover seemed implausible. The society has been and remains matriarchal and matrilocal, with husbands moving in with their wives' families. "Houses are owned by women, and women were the foundation of households," says co-author Stephanie Wynne-Jones, an archaeologist at the University of York.

Seen through that lens, the genetic results look very different—and put the "Persian princes" chronicle back into play. "Archaeologists had been debating the origins of the Swahili people," Brielle says, "and the whole time they had their own story—which, it turns out, might not be mythology."

Merchants from Persia, the authors argue, sailed south across the Arabian Sea on monsoon breezes. After landing on the Swahili coast, they married into powerful local families before setting out to sea once more. "They would stay in the places they traded, sometimes for years," Wynne-Jones says.

The unions were win-win: Local elites gained blood ties to far-off trading networks and the prestige of being related to people in Persia, an important center of the medieval Muslim world. Merchants, meanwhile, gained a foothold in local markets along with trusted partners to run their business during long overseas voyages. "It's a cunning strategy," Schmidt says. "They're buying into existing infrastructure and networks."

Based on the rate at which genes combined over generations, the team estimated the African-Persian mixing began by 1000 C.E. That timing suggests the stone towns themselves were a home-grown phenomenon. Recent excavations by Kusimba, Horton, and others have shown the coast's distinctive architecture, built of carved coral blocks, evolved from wattle-and-daub buildings beginning around 700 C.E., long before the now-documented genetic influx from abroad began. "We have 300 years of Swahili civilization preceding this," Horton says. "What we're seeing is an event where Persians arrive into a well-formed culture or civilization and very rapidly get entangled."

New waves of migration and settlement followed. The researchers saw an increase in Arabian-related DNA in the burials around 1500 C.E., as trade shifted from Persia and India to the Arabian Peninsula. India, too, was a source of migration to the region, contributing a small but measurable signature to the medieval DNA samples.



Medieval tombs in Kenya yielded DNA that revealed overseas roots of the region's trading culture.

PHOTO: CHAPURUKHA KUSIMBA/UNIVERSITY OF SOUTH FLORIDA

Today, many Swahili people have little genetic relationship with the medieval individuals in the study. Samples from almost 200 modern people who identified as Swahili showed that only those with ancestral ties to coastal towns retained large amounts of Persian ancestry. "These results highlight an important lesson," says David Reich, a geneticist at Harvard and co-author of the study. "While we can learn about the past with genetics, it does not define identity."

It is in the oral histories, often overlooked or neglected by modern scholars, that the Persian princes live on. "Sometimes we've been prone to dismiss local chronicles as made-up," Kusimba says. "Probably we ought to take oral tradition more seriously." ■

CONSERVATION POLICY

Gaps seen in China's crackdown on wildlife trade

Stricter rules could prevent disease outbreaks, but allowances for fur farming spur concern

By **Dennis Normile**

For years, scientists and conservationists have urged China's government to crack down on a thriving trade in wild animals that they say both threatens the nation's rich biodiversity and increases the risk that a dangerous disease will jump from wildlife to humans. Now, some of those pleas are being answered: On 1 May, officials will begin to enforce a strengthened Wildlife Protection Law that, together with other recent rules, expands China's list of protected species and criminalizes the sale or consumption of meat from certain animals—including raccoon dogs—known to harbor viruses that can infect humans.

Many scientists are welcoming the new law, which was finalized in December 2022. It "clearly prohibits the consumption, hunting, trade, and transport of terrestrial animals that grow and breed naturally in the wild," says Jiahai Lu, an epidemiologist at Sun Yat-Sen University, Guangzhou. Others say the restrictions could help crimp the wildlife meat trade that touched off the 2002 severe acute respiratory syndrome (SARS) outbreak and may have sparked the COVID-19 pandemic. Last month, for example, researchers released an analysis of genetic material collected at the Huanan Seafood Wholesale Market in Wuhan, China, that suggests raccoon dogs and other wildlife being sold illegally at the market were carrying SARS-CoV-2.

But the rules also have worrying weaknesses, researchers say. They permit, for example, farmers to raise raccoon dogs and other mammals for their fur—fueling concerns that the farms could promote the emergence of new human diseases, as pathogens flow among closely packed animals and their human keepers. "The continuation of the legal sale of these animals still represents a [zoonotic] risk, regardless of the intended purpose," says Alice Hughes, a conservation biologist at the University of Hong Kong.

The government has also relaxed rules governing the captive breeding of animals used in traditional Chinese medicine and as pets. Conservationists fear that will enable poachers to use farms to "launder" animals illegally caught in the wild to legal markets.

The moves represent the latest twist in China's efforts to regulate the wildlife trade. In early 2003, the government temporarily banned all wild animal sales after the outbreak of SARS, which studies indicate moved from bats to humans via palm civets, a mainstay of the meat trade. But it eased restrictions after the SARS threat faded. Then, in February 2020, shortly after COVID-19 was linked to the Huanan market, officials permanently banned the consumption of meat from wild species to "eradicate the bad habit of indiscriminate consumption of wildlife, [and] effectively prevent major public health risks," Xinhua, the official state news agency, said at the time. They cushioned the blow by paying compensation to farmers who had licenses to raise and sell the animals.

In May 2020, the Ministry of Agriculture and Rural Affairs clarified the extent of the ban, issuing a list of species that farmers can legally raise for meat, eggs, and milk. In addition to traditional livestock such as pigs and chickens, it identifies 16 "special" animals deemed to pose low human health risks. These include several species of deer, as well as animals not native to China, such as ostrich and emu.

The ministry also essentially approved the practice of farming minks, silver and arctic foxes, and raccoon dogs for fur but not meat. China is already one of the world's largest fur producers, turning out 27 million animal pelts in 2021, according to ACTAsia, an animal welfare group. Researchers fear that, without strict biosafety measures, China's fur farms could become disease hot spots. Farming creates the potential for virus spread, propagation, and transmission to humans, says veterinary virologist Conrad Freuling of the Friedrich Loeffler Institute. His research group, for example, has found that raccoon dogs can be infected with SARS-CoV-2 and transmit the virus to other animals even while showing "only subtle clinical signs" of illness.



New rules allow China's farmers to raise raccoon dogs and other mammals for their fur, but not sell their meat. Researchers fear farms could become disease hot spots.

Others point to SARS-CoV-2 outbreaks on European mink farms as an example of the dangers of fur farming. In the Netherlands in early 2020, the virus apparently spread from farm workers to minks, and then jumped from farm to farm and even back to humans. Farmers gassed 1 million minks to prevent them from becoming a reservoir for the virus. "Given the consequences as seen with COVID-19," the threat of farm-based outbreaks must "be taken seriously," Freuling says.

China is taking steps to strengthen animal disease surveillance, quarantine controls, and the use of protective equipment among farm workers, Lu says. But given the scale of China's animal farms, they are "inevitably a ticking bomb for zoonotic disease to emerge," says Ceres Kam, a wildlife campaigner with the Environmental Investigation Agency, a nongovernmental organization.

Conservationists, meanwhile, say making it easier for farmers to raise animals used in traditional Chinese medicine could put endangered pangolins, snakes, and other animals at greater risk. A major concern is that farmers will "use wild-caught individuals to boost the farmed population," says a Chinese wildlife researcher, who spoke anonymously because of the sensitivity of the topic. The change, the scientist says, is "the most widely criticized part of this law's revision." ■

With reporting by Bian Huihui.

PHOTO: BIOSPHOTO/ALAMY STOCK PHOTO

FEATURES

HORSE NATIONS

After the Spanish conquest, horses transformed Native American tribes much earlier than historians thought

By **Andrew Curry**

Scattered across the prairie east of the Colorado Front Range are rings of ancient stones. The rings were used to anchor tipis, and they measure barely 2 meters across. Matt Reed, the Pawnee Nation of Oklahoma's tribal historic preservation officer, says that tiny footprint comes as a surprise to modern Pawnee, whose traditional tipis are big enough to fit whole families.

The change, Reed explains, resulted from the introduction of the horse. For millennia, the Pawnee had relied on dogs to haul their belongings on bison hunting trips; when they acquired horses, the impact was immediate and dramatic. "They allowed us to carry more gear, pull more food, have bigger tipis," Reed says. "It's so hard to imagine our culture without horses, it boggles your mind."

For Native peoples on the Great Plains grasslands that stretch from the Rocky Mountains to the Missouri River, horses took on a central economic and military role, enabling bison hunting on a large scale and raiding across vast distances. "The introduction of this technology, of horses, changed Great Plains cultures," says Carlton Shield Chief Gover, a member of the Pawnee Nation and an archaeologist at the Indiana University Museum of Archaeology and Anthropology. "It's the equivalent of the airplane. It shrank the world." Knowing when that happened is critical, he says. A new study on p. 1316 this week in *Science*, of which Shield Chief Gover is a co-author, offers a startling answer.

Centuries ago, the Americas were apparently horseless—even though *Equus* had evolved in the Americas more than 4 million years ago, spreading west from there into Eurasia and Africa. When the ancestors of Native Americans entered North America toward the end of the last ice age, more than 14,000 years ago, they would have encountered herds of wild horses. From the archaeological evidence—cutmarks on bones found at a handful of sites—it seems early Americans hunted horses and used their bones as tools, but did not domesticate or ride them. And by 5000 years ago at the latest, the fossil record suggests, North America's horses were gone. Along with nearly 40 other species of megafauna, from saber-toothed tigers and mammoths to camels, they were wiped out by hunting, climate change, or both.

It wasn't until 1519 C.E., when Spanish conquistador Hernán Cortés made landfall on the Gulf Coast of Mexico, that horses entered the Americas again. His 16 horses stunned local people, and the shock helped him defeat the Aztec Empire just 2 years later. In the centuries that followed, the horse spread once again across the continent, this time as a status symbol, means of transport, and hunting companion rather than prey. In the process, it set off massive human migrations, as some Native groups shifted to more mobile lifestyles. It also unleashed struggles over resources on the Plains and elsewhere.

Historians have tended to date the widespread adoption of the horse by Native peoples to the 18th century, when the first European travelers recorded its presence in the central and northern Plains. But in the sweeping new study, based on archaeological evidence, radiocarbon dating, isotope analysis, and ancient DNA, Shield Chief Gover and dozens of other researchers conclude that horses had made it that far north up to a century earlier. The study shows they had begun to spread within a few decades after the Spanish introduced them to the Southwest in the 16th century.

"It's a really detailed, round, robust, multimethodology way of looking at the data set that starts to define, from an archaeological perspective, when horses appear in the American West," says University of Oxford archaeologist Peter Mitchell, who was not involved with the research. "This paper totally changes the game."

DESPITE THE HORSE'S iconic importance to so many Native cultures, little archaeological research has been done on its spread. Written records reveal such details as the names of Cortés's steeds and the first time Spanish soldiers encountered Comanche warriors on horseback. But because the horse's dispersal

PHOTO: (OPPOSITE PAGE) JACQUELYN CORDOVA



Ethnohistorian Yvette Running Horse Collin has studied the Lakotas' ancient bond with horses, reflected in ceremonial battle gear.



A healed fracture on a foal's right jaw suggests tribes in Wyoming cared for horses by the 17th century.

PHOTO: PETER BITTNER

happened mostly out of sight of European chroniclers, much of the process wasn't documented, or was written in a way that emphasized the role of Spanish and later settlers.

Based on those written sources, many historians have tended to compress the adoption of the horse by tribes throughout the Great Plains and Rocky Mountains into a pivotal half-century, beginning in 1680 with a bloody revolt against Spanish rule by Pueblo people in New Mexico and ending with the first European accounts of horses on the northern Plains. After the uprising, the story goes, the Pueblos sold thousands of horses that had belonged to the expelled Spanish to neighboring tribes. "What historians argue is that the Pueblo Revolt pushes a volume of horses, enough to transform tribes," far to the north, says Dan Flores, an emeritus historian at the University of Montana, Missoula.

"In the aftermath of the Great Southwestern Rebellion, the horse frontier moved rapidly outward from New Mexico along the ancient Indigenous trade routes," Oxford historian Pekka Hämäläinen writes in *Indigenous Continent: The Epic Contest for North America*, a history of Native societies published last year. "A Rocky Mountain trade chain had carried horses to the Northwestern Plains by the 1730s. … The horse trade ignited a technological revolution that reconfigured several Indigenous worlds within a generation."

Historians weren't fazed by the implication that it took so little time for the horse to be incorporated into Native economies, military strategy, and religious ritual across a vast geographic area. Flores and others argue that adoption was facilitated by contact with the Spanish, who employed Pueblo people to herd horses and other livestock in New Mexico in the early 1600s. "You have to have a horse-riding culture and technology, along with the animal, to make this work," he says.

Native accounts contradicted the timeline centered on the Pueblo Revolt, suggesting some tribes had acquired horses much earlier, but "oral tradition was discounted," says Comanche historian Jimmy Arterberry, a co-author of the *Science* study. "The end result has been to discredit the antiquity of the relationship between Native people and horses," adds University of Colorado, Boulder, archaeologist William Taylor, also a co-author.

Until recently, archaeologists took the compressed historical timeline for granted. When they excavated horse remains on the Great Plains, they usually assumed the bones were either very old, dating from before the disappearance of horses many millennia ago, or very recent, from animals brought to the Plains by European settlers. As a

result, many horse remains found on the Great Plains wound up in paleontological collections rather than archaeology labs. Just a handful had been radiocarbon dated. The *Science* study includes dozens of new dates and shifts the timeline earlier.

Dates of horse remains from sites in Wyoming and Nebraska, for example, show people far beyond the Spanish frontier were breeding, feeding, herding, and caring for horses—and probably riding them—beginning sometime after 1550, and had thoroughly incorporated them into their societies by 1650 at the latest. That "provides more time for the transformations the horse brought about to take place," Mitchell says.

Some historians say they've made steps in that direction, using ethnographic and linguistic evidence. The idea that horses might have been adopted by Native groups in New Mexico and beyond in the 1600s, prior to the Pueblo Revolt, "is pretty much in line with what historians have been writing for the last 20 or 30 years," says Ted Binnema, a historian at the University of Northern British Columbia. "The entire military history of the Plains seems to make sense with this timeline."

But pushing the arrival of the horse back much further is a stretch, he says. Spanish troops entering New Mexico in 1598 make no mention of encountering mounted warriors. "Any evidence Indigenous groups were already equestrian before 1598 would be a big discovery," Binnema says. "But this isn't enough to convince me."

EMILY LENA JONES keeps what the zooarchaeologist at the University of New Mexico, Albuquerque, considers the key evidence in a clear plastic organizing tub. Inside are Ziploc bags filled with fragmented, bleached-white horse bones from a village on the outskirts of Albuquerque. The site, called Paa'ko, was settled by Puebloan people beginning in 1525, and briefly used as a mission by the Spanish between 1650 and the Pueblo Revolt of 1680.

For decades after it was first excavated in the 1930s, researchers assumed the handfuls of horse bones at Paa'ko dated from that short-lived Spanish occupation. They identified them as the remains of the Spanish friar's mounts, eaten by the Native residents.

When Jones radiocarbon dated the bones in 2020, she was surprised to find they were at least 400 years old and probably even predated the establishment of the first Spanish settlement in New Mexico in 1598. Given Spanish records describing the horse as a tightly restricted military asset, the idea that they were in areas not under Spanish control came as a surprise. The new timing, she says, "opens up a wide range of cultural change happening outside of European view." Jones thinks Paa'ko might have been an early transit point where Native people moved horses from Spanish-controlled areas east to the Plains.

As she, Taylor, and other members of their team gathered radiocarbon dates from sites across the Great Plains, they found more samples that predated the Pueblo Revolt—some by nearly a century. At a site in southwestern Wyoming called Blacks Fork, a young foal was buried together with three coyote skulls, evidence it may have been part of a religious ceremony. Analysis of the animal's bones and teeth shows it was about 6 months old when it died. Bone formations on its skull suggested it was tethered, and a healed facial fracture showed it had been kicked by another horse—all support for the idea that it was kept close to other horses, and given veterinary care to help it recover from its wound.

Some researchers who analyzed the site in the 1990s discounted the possibility that Native people would have known how to handle a horse that way. Instead, they speculated the Blacks Fork horse might have been brought to the region by an unrecorded Spanish expedition.

But the new radiocarbon date for the Blacks Fork horse shows it was raised, tethered, and buried sometime between 1600 and 1650, many hundreds of kilometers north of Paa'ko and Spanish outposts in New Mexico. "I was very surprised by the Blacks Fork horse. I thought we'd find some horses that were earlier than the overarching narrative said, but I didn't expect this line of inquiry to bear fruit so quickly," Jones says. "Every time we've thrown something in there, it's come back in this earlier date range."

"People are going to have to go back into their collections and start redating horses," Shield Chief Gover says. "This is upending the status quo."



Archaeologist William Taylor inspects the jawbone of a foal from Blacks Fork, Wyoming.

PHOTO: PETER BITTNER

IN HIS RED PICKUP, Taylor and his team continue to visit local and regional museums from Wyoming to Kansas in search of bones to analyze. "To engage with this material requires putting a lot of miles on the odometer and working with a lot of small collections," Taylor says. As they persuade curators to part with bits of bone for radiocarbon and chemical testing, the team also digitizes the bones with a hand-held 3D scanner. Back at their computers in Boulder, they can measure and analyze the scans digitally, or even print out plastic copies to share with the public.

The results enable them to go beyond the age of the bones to the lives of the horses—and of the humans who cared for and relied on them. In a darkened lab at the museum, Taylor and Lakota grad student Chance Ward, a member of the Cheyenne River



This Nokota mare is descended from horses that belonged to Sitting Bull, the 19th century Lakota leader who fought to defend a way of life that horses made possible.

Sioux Tribe, deploy beetles to scour the flesh from modern horse cadavers. The modern bones serve as references: Draft horses, wild horses, horses ridden in the deserts of New Mexico and Arizona or the steppes of Mongolia—all show distinctive skeletal changes that can be compared with marks on the ancient North American bones.

Cradling a 3D printed copy of a horse skull from Mongolia, Taylor points to a groove midway along the long, sloping nose, caused by pressure from a bridle. "If you press on this continuously for 10 years, you get marks that are diagnostic," he says. "It's negative space, but in 3D models we can measure it."

Similar marks in bones from early Native American horses can help prove they carried a rider, wore a bridle, or pulled a travois—a frame, made of two poles and a net, that Plains people used to drag loads. Metal ring bits adopted from the Spanish caused distinctive fractures in an animal's teeth and palate; rawhide bridles developed by Native people wore telltale grooves in the lower jaw. "Every little aspect of human activity and relationship to horses leaves a signature, if we can find it," Taylor says. "It's not just whether people rode them, but also how."

The bones contain other clues, too. Horse teeth found on the banks of the Kansas River in northeastern Kansas, and dated to before 1650—at least 30 years before the Pueblo Revolt—are a good example. Isotopes in the teeth reflect those in the groundwater where the animal was reared. They show the horse, about 9 years old when it died, had spent time farther north, perhaps in Iowa, before it was moved into Kansas. The isotopes also show it was fed corn, a wintertime staple for Plains people. "It's an incredible snapshot of an animal that was deeply integrated into Indigenous culture," Taylor says—and another bit of evidence, he thinks, that the spread of horses across the continent didn't "have anything to do with European people other than that first horse off the boat."

Instead, that rapid spread highlights the intricate trade networks and political alliances that knit together tribes from the arid Southwest to the forests of the Missouri River valley. And ultimately, understanding the movements of horses through isotopic and genetic analyses could help trace human migration during a tumultuous period. "Knowing when horses are where they are, and who has them, can add to our understanding of tribal history and politics," Shield Chief Gover says. "Archaeological research can provide tangible history that ties people back to places."

PERHAPS THE MOST striking aspect of the *Science* study is how it shows Native views and laboratory science can enhance—and also challenge—each other.

In 2018, University of Toulouse geneticist Ludovic Orlando received an email with an unusual request. Orlando had just published a wide-ranging paper tracing the lineage of horses across the world today to the steppes of Eurasia around 2000 B.C.E. The email came from ethnohistorian Yvette Running Horse Collin of the Oglala Lakota Nation, who had just defended a doctoral

PHOTO: MARY KATHERINE MORRIS PHOTOGRAPHY/SACRED WAY SAN

dissertation with a very different perspective. Drawing on Native oral histories—including the Lakota tradition that the tribe has had a relationship with the horse "since time immemorial"—Collin argued that the horse never went extinct in the Americas at all. The arrival of Spanish horses in 1519, in her view, had been a reunion rather than a reintroduction.

Intrigued, Orlando invited her to Toulouse—and eventually to collaborate on a genetic investigation into the origins of Native horses that became part of the *Science* paper. "I very rapidly realized there was room for questions about the origins of the horse," Orlando says, "and we developed an experimental design to test all this." Collin, for her part, says the horse seemed like an ideal vehicle "to further a discussion" on marrying Native and Western approaches to knowledge. "I did not come to Ludovic to help me to prove what we already know," she says.

Together with Taylor and other colleagues, the Toulouse team collected DNA from the bones of 29 horses of the historic period, from the 17th century and later, along with samples from modern horses cared for by the Lakota and other tribes. They then compared the DNA with Orlando's database of modern and ancient horse genomes.

The results showed horses on the Great Plains in the historic period were closely related to horses in Spain at the same time. By the 1770s, however, genetic signatures from British breeds began to filter into the region, and horses there today show a mix of both Spanish and British ancestry (and no link to horses the Vikings are known to have taken as far as Greenland). The researchers looked for traces of DNA hailing from the horses that had lived in the Americas in the ice age or right after. They found none.

"We did their genomes, and they look Western in origin," Orlando says. "Does that change the fact that the Lakota view the horses there today as theirs? No." Nor should it, he adds: "I don't want the world to conclude this is a genetic demonstration they had no relationship with the horse."

Unusually for a scientific paper, this one acknowledges that some interpretive gaps among the authors remained unbridged—and it includes a statement from Lakota elders, several of whom are co-authors. "Horses have been part of us since long before other cultures came to our lands," Lakota Chief Joe American Horse writes, "and we are a part of them."

The paper leaves open a tantalizing possibility. DNA recovered from soil in the Arctic suggests horses might have survived until at least 5000 years ago in parts of North America, where people hunted them and fashioned their bones into tools. Perhaps the memory of that early relationship survived for millennia and is preserved in the oral tradition of the Lakota and other groups—who then reestablished a connection with domesticated horses in the past few centuries. "It would be crazy to dismiss this idea without testing it further," Taylor says. "And now we can start to do that." More radiocarbon dates and DNA, along with other methods, might even document an intersection between ancient horses and the Lakota and other groups, he says.

Other tribes have very different oral traditions, some of which the new research reinforces. In Pawnee, Shield Chief Gover points out, the word for "horse" translates as "new dog." Other Indigenous languages, too, reflect an initial unfamiliarity with the beasts: Blackfeet called them "elk dogs," Comanche "magic dogs," the Assiniboine "great dogs." "Even in language, it shows up as 'what is this?!'" Shield Chief Gover laughs. "Our oral traditions do not say we've always had horses. This is another piece of evidence that shows oral traditions were always correct, and archaeology's catching up."

One traditional Pawnee song, still sung by elders in the tribe today, tells of a long-ago encounter with a group of mounted outsiders in armor. "The song talks about people with metal, very foreign, who started a fight. We finished it, and took horses from them," says Reed, who was not involved in the *Science* study. In 1540, as it happens, a Spanish conquistador named Francisco Vázquez de Coronado led a large expedition, including several hundred horses, north through New Mexico as far as what is now Kansas, where the Pawnee lived before they were pushed onto Oklahoma reservations. Based on Spanish accounts, the 1540 date had seemed too early for the Pawnee to have acquired horses—but the new study makes it conceivable. "This research verifies and solidifies our history, and that's important," Reed says.

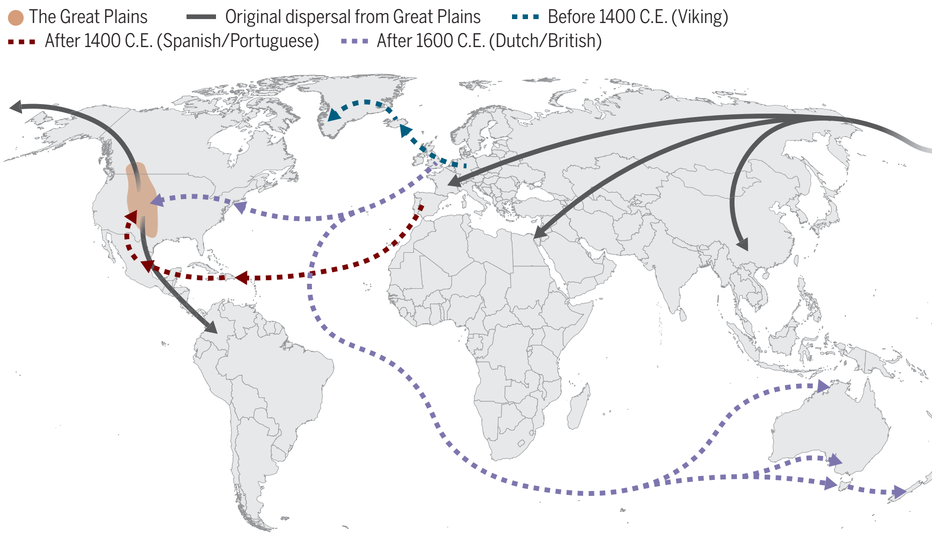
Linguistic evidence reflects how rapidly and thoroughly the horse was incorporated into Native societies. When ethnographers wrote down Pawnee, Comanche, and other Native languages in the 1800s, the vocabularies included dozens of Indigenous terms for horse anatomy, tack, appearance, and breeding, along with extensive catalogs of plants used for equine veterinary care. Just a few centuries earlier, "not only did people have no experience with horses, they had no experience with any large animals," at least domesticated ones, says Greger Larson, an Oxford geneticist who was not involved with the study. "It's a real demonstration of the plasticity of humans."

The flip side of the story is the speed with which the relationship was—almost—shattered once more. In the late 1800s, Plains tribes were stripped of their land and horses and prohibited from speaking their languages. Today the Pawnee, Comanche, and others are more likely to drive pickups than herd ponies.

To some co-authors of the new study, the results represent an opportunity to undo some of that loss. "Archaeological research can provide tangible history," Shield Chief Gover says. "Horses are history that can be touched, which is uncommon for Indigenous people today." ■

Coming home

Horses evolved millions of years ago in North America and, after spreading to Eurasia and Africa, went extinct in their homeland at the end of the last ice age. Spanish and British colonizers brought them back.



CREDITS: (GRAPHIC) D. AN-PHAM/*SCIENCE*; (DATA) W. T. TAYLOR/UNIVERSITY OF COLORADO MUSEUM OF NATURAL HISTORY

Check for updates

PERSPECTIVES

HYPOTHESIS

The link between obesity and autoimmunity

Overnutrition could lead to loss of self-tolerance by impinging on immune regulation

By **Giuseppe Matarese**[1,2]

Compelling epidemiological evidence reveals a strong association between being overweight or obese and the risk of developing autoimmune diseases (*1*). From an immunological standpoint, the cellular and molecular mechanisms linked to this association include the overstimulation of T lymphocytes by nutrient- and energy-sensing pathways. The immunometabolic state of an individual is central to the modulation of immunological self-tolerance that suppresses self-reactivity to avoid autoimmunity. Adipose tissue is an immunologically active organ that influences systemic immune responses through the production of adipocytokines, and, in turn, immune cells affect adipocyte homeostasis and metabolism through the production of pro- and anti-inflammatory cytokines (*2*). This implies that metabolic overload from obesity can affect immunometabolism, which can alter susceptibility to autoimmune diseases.

Immunological adaptations occur in response to nutritional status: Undernutrition impairs immunity, causing inefficient responses to infections and vaccinations. Conversely, overnutrition favors chronic activation of both innate and adaptive immune cells, with subsequent (low-grade) systemic inflammation. These phenomena occur through the engagement of intracellular nutrient- and energy-sensing pathways and the NACHT, LRR and PYD domains–containing protein 3 (NLRP3) inflammasome, which is a sensor of metabolic stress that is induced by an excess of glucose and lipids, especially in macrophages (*2*, *3*).

Obesity is a risk factor for autoimmune conditions such as type 1 diabetes (T1D) and multiple sclerosis (MS) (*4*, *5*). Environmental and lifestyle factors that increase MS risk include smoking, sun exposure, low vitamin D, Epstein-Barr virus infection, and high body mass index (BMI). Prospective longitudinal studies in young obese individuals found a 1.6- to 1.9-fold increase in the risk of developing MS during adolescence and young adulthood (but not at the time of MS onset); this association with obesity was also confirmed in carriers of the human leukocyte antigen (HLA)–DRB1*15:01 susceptibility allele that is responsible for the presentation of myelin self-antigens to autoreactive T cells (*5*). Similarly, higher BMI at birth is associated with higher T1D susceptibility in children. Indeed, the incidence of T1D increased almost linearly with a higher birth weight (1.7% increase in incidence per 100-g increase in birth weight) (*4*).

Mechanistically, it has been suggested that increased body adiposity promotes the hyperactivation of intracellular nutrient- and energy-sensing pathways [such as mechanistic target of rapamycin (mTOR)] with subsequent metabolic overload in peripheral tissues, including resident immune cells that are involved in both effector and regulatory immune responses (*6*). For example, in obese naïve-to-treatment MS patients, the adipocytokine leptin (secreted in proportion to BMI to inhibit food intake), together with elevated amounts of circulating nutrients, was found to boost inflammatory immune responses. High levels of leptin and nutrients cause constitutive overactivation of mTOR in T cells, with subsequent dysregulated T cell receptor (TCR)–mediated signaling. Overactive mTOR in T cells mimics a strong, supra-physiological TCR stimulation that is not permissive for transcription of the forkhead-box P3 (*FOXP3*) gene, the expression of which is pivotal for the induction and maintenance of anti-inflammatory $CD4^+CD25^+FOXP3^+$ regulatory T cells (T_{regs}) (*2*, *6*). Through leptin overproduction, obesity impairs the proliferation of anti-inflammatory thymic T_{regs} and their peripheral differentiation from $CD4^+CD25^-$ conventional T (T_{conv}) cell precursors (*7*). Obesity also promotes conversion of T_{conv} cells into pathogenic inflammatory T helper 1 (T_H1) and T_H17 cells, thus increasing the risk of altered immunological self-tolerance (see the figure).

Overall, nutrient- and leptin-induced mTOR overactivation inhibits peripheral T_{reg} proliferation and suppressive function and enhances obesity-associated T_H1 and T_H17 cell differentiation, with a higher risk of MS-associated myelin damage (*2*, *7*). Further, a recent report demonstrated that obese mice converted the classical T_H2-predominant immune responses of atopic dermatitis into a severe disease predominantly characterized by T_H17-driven inflammation that was caused by reduced activity of the peroxisome proliferator-activated receptor-γ (PPAR-γ) transcription factor (*8*). The expression of PPAR-γ in adipose tissue was also necessary for the development and function of adipose-tissue resident T_{regs}, suggesting a further bidirectional link between adipose tissue biology and immune tolerance that involves T_{regs} (*2*, *7*).

Physiological nutrients and leptin fluctuations due to daily cycles of fasting and feeding determine oscillations in mTOR activity that are lost in obesity because of excessive food intake. Therefore, in individuals with a normal BMI and physiological cycles of feeding and fasting, the maintenance and perpetuation of self-tolerance are associated with oscillations of mTOR activity in T_{regs}. This appears to be necessary for T_{reg} expansion and function in sufficient numbers to suppress pathogenic T_H1 and T_H17 cells and thus autoimmunity (*2*, *6*, *7*).

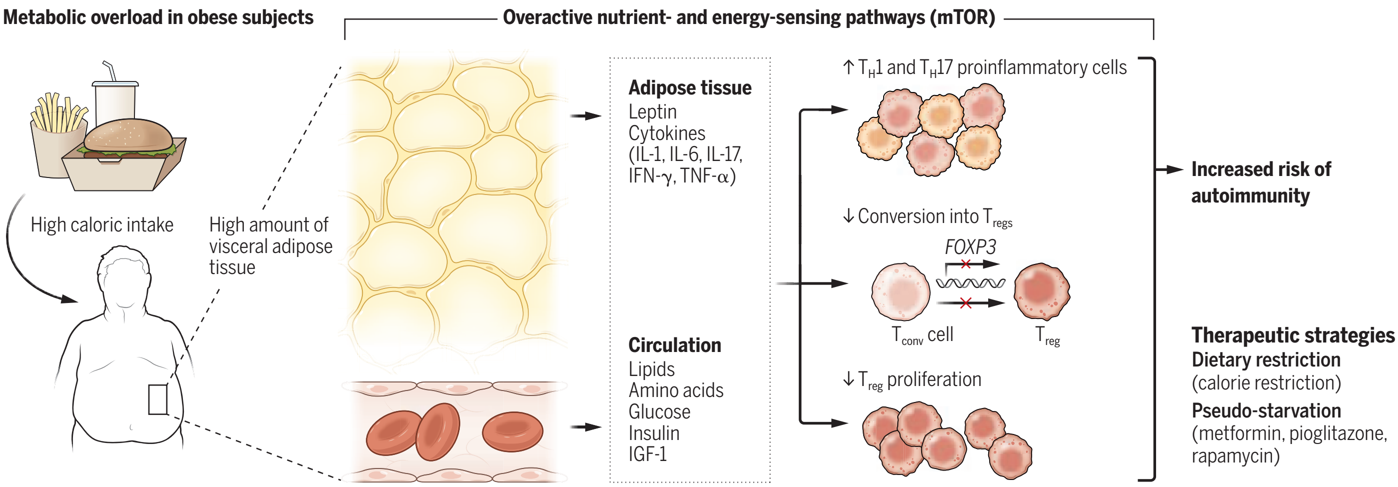
Of note, mTOR represents a key intracellular node at the crossroad of amino acid, glucose, and lipid metabolism. Furthermore, growth factors linked to nutrition and metabolism, such as leptin, insulin, and insulin-like growth factor 1 (IGF-1), activate mTOR signaling in immune cells, which affects systemic and intracellular immunometabolism and thus inflammation and autoimmunity (*2*, *6*). Adipose tissue also secretes inflammatory cytokines such as interleukin-1 (IL-1), tumor necrosis factor–α (TNF-α), IL-6, IL-17, and interferon-γ (IFN-γ), as well as leptin, which leads to a higher susceptibility to peripheral tissue damage and autoimmunity. Therefore, I propose that metabolic workload—induced by nutrients, adipocyte-derived growth factors, and adipocytokines—may represent an accelerator of autoimmune disorders in people who typically consume an obesogenic Western diet.

It has been demonstrated in mice and in

[1]Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università di Napoli Federico II, Napoli, Italy. [2]Istituto per l'Endocrinologia e l'Oncologia Sperimentale, Consiglio Nazionale delle Ricerche (IEOS-CNR), Napoli, Italy. Email: giuseppe.matarese@unina.it

Altered immunometabolism in obesity could lead to autoimmunity

In obese people, excessive stimulation of immune cell nutrient- and energy-sensing pathways (such as increased mTOR activity) through circulating and adipose tissue–derived factors can skew immune cell differentiation. This increases numbers of proinflammatory T_H1 and T_H17 cells and decreases Tregs, owing to reduced conversion of T_{conv} cells and reduced T_{reg} proliferation, which increase the risk of losing self-tolerance. Immunometabolic strategies may help to restore homeostasis.



FOXP3, forkhead-box P3; IFN-γ, interferon-γ; IGF-1, insulin-like growth factor 1; IL, interleukin; mTOR, mechanistic target of rapamycin; T_{conv}, conventional T; T_H, T helper; TNF-α, tumor necrosis factor–α; T_{reg}, regulatory T cell.

humans that adaptive and innate immune cells can directly influence the pathophysiological events that lead to obesity and obesity-associated metabolic abnormalities (*2*). This could also contribute to the reduction in T_{reg} numbers observed in obese people. There is an anatomical and functional cross-talk between adipose tissue and the immune system. Indeed, both primary lymphoid organs (bone marrow and thymus) and secondary lymphoid organs (lymph nodes) are generally embedded in and surrounded by adipose tissue. This contiguity allows T cells, T_{regs}, B cells, dendritic cells, and macrophages to home to adipose tissue. Additionally, adipocytes can express immune-like behaviors (*2*). For example, adipocytes can clear intracellular bacteria using the same nuclear-binding oligomerization domain 1 (NOD1) pathogen-sensing system of innate immune cells (*9*). Changes in T_{reg} numbers and function observed in obesity may also affect susceptibility to infections and cancer (*2*). Indeed, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is associated with the production of autoantibodies and is more severe in obese individuals (*10*). Additionally, cancer immunotherapy responses are better in obese people than in patients with a lower BMI (*2*).

Polygenic obesity (predisposition to obesity caused by multiple genetic variants and environmental factors) has also been proposed to be an autoimmune-like disease, whereby T cells respond to unknown adipocyte antigens and trigger subsequent uncontrolled food intake, although the mechanism remains to be fully elucidated (*11*). It is notable that $CD4^+$ T cells isolated from obese mice could transfer an "obesity memory" by promoting weight gain when injected into normal-weight, immune-deficient recipients (*11*). Thus, it appears that obesity associates with a higher susceptibility to develop autoimmunity not only because adipose tissue boosts autoinflammatory responses but also because obesity itself has autoimmune-like features.

A promising possibility is the manipulation of immune tolerance and autoimmunity through immunometabolic interventions: reduced food and/or calorie intake. Although the idea that fasting could modulate immune responses and alleviate symptoms of autoimmune diseases had mostly been dismissed, studies over the past 20 years have provided evidence that supports the therapeutic potential of behavioral changes and nutritional strategies such as diet, caloric restriction (CR), and different fasting regimens (*2*, *12*). Mild CR, intermittent fasting, and a ketogenic diet have each shown beneficial effects in mouse models of autoimmunity, including experimental autoimmune encephalomyelitis (EAE), experimental rheumatoid arthritis, and experimental colitis (*2*, *12*). I suggest that "starving" pathogenic inflammatory T_H1 and T_H17 cells could lead to better control of local and systemic inflammation. Similarly, CR allows expansion of T_{regs} in mice and humans by promoting their generation, proliferation, and function, thereby controlling autoimmunity (*2*, *7*, *12*).

Because adherence to dietary changes is not always possible, a proposed alternative approach is "pseudo-starvation," whereby drugs that regulate immunometabolism mimic fasting (*13*). A prototypical example is the mTOR inhibitor rapamycin. Additionally, metformin, an activator of AMP-activated protein kinase (AMPK) that is used to treat type 2 diabetes and overweight individuals, not only controls glucose tolerance but also has anti-inflammatory actions through AMPK-mediated mTOR inhibition (*13*). Metformin attenuated EAE induction by restricting the infiltration of mononuclear cells into the central nervous system (CNS) and down-regulating the expression of inflammatory cytokines, inducible NO synthase, cell-adhesion molecules, matrix metalloproteinase-9, and chemokines in T_H17 cells (*14*). These effects were also observed in a study of MS patients with metabolic syndrome (*15*).

First-line drug treatments for MS (either IFN-β or glatiramer acetate) in combination with metformin provided a statistically significant improvement of disease and reduced CNS lesions. These effects were associated with lowered circulating leptin and T_H1 and T_H17 inflammatory cytokines and increased numbers of peripheral T_{regs} (*15*). Similarly, pioglitazone, an activator of PPAR-γ with antidiabetic effects, also provided a metabolic signal of pseudo-starvation to immune cells from treated MS patients by increasing insulin sensitivity and reducing circulating glucose and leptin levels (*15*). In EAE, pioglitazone treatment controlled the disease course with reduced CNS infiltrates and decreased inflammatory cytokine production and T_H1 and T_H17 differentiation (*15*).

GRAPHIC: A. FISHER/*SCIENCE*

Also, it is interesting to note that classical anti-inflammatory and immunosuppressive drugs such as salicylate and methotrexate can convey metabolic signals of pseudo-starvation to immune cells through the activation of AMPK (*13*), along with their classical mechanisms of action. Overall, conferring metabolic signals of pseudo-starvation could be valuable in down-regulating autoinflammatory responses.

It is remarkable that during CR, T cells reprogram their transcriptional signature toward anti-inflammatory properties that limit tissue damage and prolong life span in mice and humans (*7*, *12*). Also, CR induces extensive adaptations in the gut microbiota toward the production of anti-inflammatory metabolites that affect local and systemic immunometabolism (*12*). Molecules that interact with adipocyte-derived leptin can modulate immune function in various ways depending on metabolic status. For example, neuroendocrine mediators with appetite-stimulating activity such as ghrelin and neuropeptide Y have opposite effects from those of leptin, not only on satiety but also on the peripheral immune responses because they block the secretion of T_H1 and T_H17 cytokines and suppress EAE (*2*). Remaining areas for study include the molecular dissection of how single nutrients (i.e., lipids, carbohydrates, and proteins) affect immunological self-tolerance and the temporal window in which CR is an effective therapeutic regimen for obesity-associated autoimmunity. ■

REFERENCES AND NOTES

1. A. Lerner, P. Jeremias, T. Matthias, *Int. J. Celiac. Dis* **3**, 151 (2015).
2. P. de Candia *et al.*, *J. Exp. Med*. **3**, 218 (2021).
3. A. Christ *et al.*, *Cell* **172**, 162 (2018).
4. C. A. March, D. J. Becker, I. M. Libman, *Front. Endocrinol*. **22**, 12 (2021).
5. A. K. Hedström *et al.*, *Neurol. Neuroimmunol. Neuroinflamm*. **8**, e912 (2020).
6. M. H. Do *et al.*, *J. Exp. Med*. **217**, e20190848 (2020).
7. P. de Candia, C. Procaccini, C. Russo, M. T. Lepore, G. Matarese, *Immunity* **55**, 1981 (2022).
8. S. P. Bapat *et al.*, *Nature* **604**, 337 (2022).
9. G. Caputa *et al.*, *Cell Metab*. **34**, 747 (2022).
10. E. Y. Wang *et al.*, *Nature* **595**, 283 (2021).
11. J. Zou *et al.*, *Cell. Mol. Immunol*. **15**, 630 (2018).
12. O. Spadaro *et al.*, *Science* **375**, 671 (2022).
13. L. A. O'Neill, D. G. Hardie, *Nature* **493**, 346 (2013).
14. Y. Sun *et al.*, *J. Neuroimmunol*. **292**, 58 (2016).
15. L. Negrotto, M. F. Farez, J. Correale, *JAMA Neurol*. **73**, 520 (2016).

ACKNOWLEDGMENTS

G.M. thanks P. de Candia and A. La Cava for critically reading the manuscript and S. Bruzzaniti for assistance with the figure. G.M. is funded by Fondazione Italiana Sclerosi Multipla (FISM grant 2018/S/5), Progetti di Rilevante Interesse Nazionale (PRIN grant 2017 K55HLC 001), Ministero della Salute (grant RF-2019-12371111), and Ministero dell'Università e Ricerca (INF-ACT grant PE00000007). This work is dedicated to the memory of S. Zappacosta and E. Papa.

10.1126/science.ade0113

NEURODEGENERATION

Effective immunotherapy occurs in neurons

Neuronal clearance of proteins in the brain proves to be important for immunotherapy

By **Rebecca M. Nisbet**

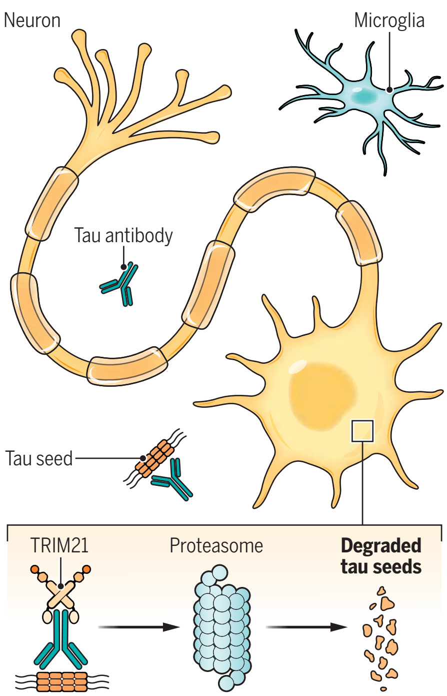
The main pathological hallmark of a group of neurodegenerative diseases called tauopathies is the formation of intracellular aggregates composed of the tau protein in the brain. Despite promising results in preclinical studies, tau immunotherapies in clinical development for the treatment of tauopathies, including Alzheimer's disease, have thus far failed to improve patient cognition. Owing to its critical pathological role, most still argue that tau is an excellent therapeutic target. Therefore, better understanding of the mechanisms by which tau antibodies can remove pathogenic tau from the brain and how these processes can be exploited is paramount for the design of second-generation immunotherapies. On page 1336 of this issue, Mukadam *et al.* (*1*) show that the cytoplasmic antibody receptor and E3 ubiquitin ligase tripartite motif-containing 21 (TRIM21) is required for effective tau immunotherapy in a tauopathy mouse model, providing an area of focus for the development of future tau antibodies.

Tau, a microtubule-associated protein, plays a central role in the stabilization of microtubules. In disease, however, tau becomes hyperphosphorylated, misfolds, and self-aggregates into insoluble inclusions within neurons and glia. This process ultimately leads to neuron degeneration by disrupting normal neuronal activity. Current immunotherapeutics that target tau are based on the hypothesis that misfolded pathogenic tau is released extracellularly and taken up by neighboring neurons. The internalized pathogenic tau, called a tau seed, facilitates the recruitment and misfolding of the naive monomeric tau in the recipient neuron, resulting in the formation of toxic insoluble aggregates. This process of tau seeding is thought to be the mechanism by which tau pathology spreads from one brain region to another in disease (*2*).

Tau antibodies are predicted to engage extracellular tau seeds and either block seed entry into neurons, promote neuronal uptake and clearance, or promote microglia clearance of the engaged tau. Microglia are macrophage-like immune cells in the central nervous system that display antibody receptors on their surface. The engagement of an antibody-antigen complex by the surface receptor results in microglial phagocytosis and clearance of the complex. Unexpectedly, immunotherapeutic treatment of a tauopathy mouse model with modified tau antibodies that exhibit reduced binding (*3*) or no binding (*4*) to microglial antibody receptors demonstrated a superior

Antibody-mediated clearance of tau seeds

Tau antibodies can engage pathogenic tau seeds in the extracellular environment. Once engaged, the antibody-seed complex is internalized by neurons where it can bind tripartite motif-containing 21 (TRIM21) in the cytoplasm. TRIM21 then stimulates degradation of the tau-antibody complex with the involvement of the ubiquitin-proteasome system. This clearance of the tau-antibody complex prevents tau seeding and reduces pathological tau accumulation in mice.



GRAPHIC: A. MASTIN/*SCIENCE*

Also, it is interesting to note that classical anti-inflammatory and immunosuppressive drugs such as salicylate and methotrexate can convey metabolic signals of pseudo-starvation to immune cells through the activation of AMPK (*13*), along with their classical mechanisms of action. Overall, conferring metabolic signals of pseudo-starvation could be valuable in down-regulating autoinflammatory responses.

It is remarkable that during CR, T cells reprogram their transcriptional signature toward anti-inflammatory properties that limit tissue damage and prolong life span in mice and humans (*7*, *12*). Also, CR induces extensive adaptations in the gut microbiota toward the production of anti-inflammatory metabolites that affect local and systemic immunometabolism (*12*). Molecules that interact with adipocyte-derived leptin can modulate immune function in various ways depending on metabolic status. For example, neuroendocrine mediators with appetite-stimulating activity such as ghrelin and neuropeptide Y have opposite effects from those of leptin, not only on satiety but also on the peripheral immune responses because they block the secretion of T_H1 and T_H17 cytokines and suppress EAE (*2*). Remaining areas for study include the molecular dissection of how single nutrients (i.e., lipids, carbohydrates, and proteins) affect immunological self-tolerance and the temporal window in which CR is an effective therapeutic regimen for obesity-associated autoimmunity. ■

REFERENCES AND NOTES

1. A. Lerner, P. Jeremias, T. Matthias, *Int. J. Celiac. Dis* **3**, 151 (2015).
2. P. de Candia *et al.*, *J. Exp. Med.* **3**, 218 (2021).
3. A. Christ *et al.*, *Cell* **172**, 162 (2018).
4. C. A. March, D. J. Becker, I. M. Libman, *Front. Endocrinol.* **22**, 12 (2021).
5. A. K. Hedström *et al.*, *Neurol. Neuroimmunol. Neuroinflamm.* **8**, e912 (2020).
6. M. H. Do *et al.*, *J. Exp. Med.* **217**, e20190848 (2020).
7. P. de Candia, C. Procaccini, C. Russo, M. T. Lepore, G. Matarese, *Immunity* **55**, 1981 (2022).
8. S. P. Bapat *et al.*, *Nature* **604**, 337 (2022).
9. G. Caputa *et al.*, *Cell Metab.* **34**, 747 (2022).
10. E. Y. Wang *et al.*, *Nature* **595**, 283 (2021).
11. J. Zou *et al.*, *Cell. Mol. Immunol.* **15**, 630 (2018).
12. O. Spadaro *et al.*, *Science* **375**, 671 (2022).
13. L. A. O'Neill, D. G. Hardie, *Nature* **493**, 346 (2013).
14. Y. Sun *et al.*, *J. Neuroimmunol.* **292**, 58 (2016).
15. L. Negrotto, M. F. Farez, J. Correale, *JAMA Neurol.* **73**, 520 (2016).

ACKNOWLEDGMENTS

G.M. thanks P. de Candia and A. La Cava for critically reading the manuscript and S. Bruzzaniti for assistance with the figure. G.M. is funded by Fondazione Italiana Sclerosi Multipla (FISM grant 2018/S/5), Progetti di Rilevante Interesse Nazionale (PRIN grant 2017 K55HLC 001), Ministero della Salute (grant RF-2019-12371111), and Ministero dell'Università e Ricerca (INF-ACT grant PE00000007). This work is dedicated to the memory of S. Zappacosta and E. Papa.

10.1126/science.ade0113

NEURODEGENERATION

Effective immunotherapy occurs in neurons

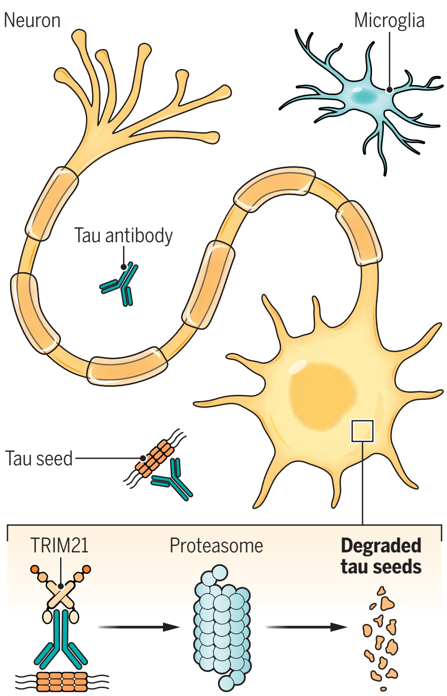
Neuronal clearance of proteins in the brain proves to be important for immunotherapy

By **Rebecca M. Nisbet**

The main pathological hallmark of a group of neurodegenerative diseases called tauopathies is the formation of intracellular aggregates composed of the tau protein in the brain. Despite promising results in preclinical studies, tau immunotherapies in clinical development for the treatment of tauopathies, including Alzheimer's disease, have thus far failed to improve patient cognition. Owing to its critical pathological role, most still argue that tau is an excellent therapeutic target. Therefore, better understanding of the mechanisms by which tau antibodies can remove pathogenic tau from the brain and how these processes can be exploited is paramount for the design of second-generation immunotherapies. On page 1336 of this issue, Mukadam *et al.* (*1*) show that the cytoplasmic antibody receptor and E3 ubiquitin ligase tripartite motif-containing 21 (TRIM21) is required for effective tau immunotherapy in a tauopathy mouse model, providing an area of focus for the development of future tau antibodies.

Antibody-mediated clearance of tau seeds

Tau antibodies can engage pathogenic tau seeds in the extracellular environment. Once engaged, the antibody-seed complex is internalized by neurons where it can bind tripartite motif-containing 21 (TRIM21) in the cytoplasm. TRIM21 then stimulates degradation of the tau-antibody complex with the involvement of the ubiquitin-proteasome system. This clearance of the tau-antibody complex prevents tau seeding and reduces pathological tau accumulation in mice.



GRAPHIC: A. MASTIN/*SCIENCE*

Tau, a microtubule-associated protein, plays a central role in the stabilization of microtubules. In disease, however, tau becomes hyperphosphorylated, misfolds, and self-aggregates into insoluble inclusions within neurons and glia. This process ultimately leads to neuron degeneration by disrupting normal neuronal activity. Current immunotherapeutics that target tau are based on the hypothesis that misfolded pathogenic tau is released extracellularly and taken up by neighboring neurons. The internalized pathogenic tau, called a tau seed, facilitates the recruitment and misfolding of the naive monomeric tau in the recipient neuron, resulting in the formation of toxic insoluble aggregates. This process of tau seeding is thought to be the mechanism by which tau pathology spreads from one brain region to another in disease (*2*).

Tau antibodies are predicted to engage extracellular tau seeds and either block seed entry into neurons, promote neuronal uptake and clearance, or promote microglia clearance of the engaged tau. Microglia are macrophage-like immune cells in the central nervous system that display antibody receptors on their surface. The engagement of an antibody-antigen complex by the surface receptor results in microglial phagocytosis and clearance of the complex. Unexpectedly, immunotherapeutic treatment of a tauopathy mouse model with modified tau antibodies that exhibit reduced binding (*3*) or no binding (*4*) to microglial antibody receptors demonstrated a superior

ability to reduce tau pathology compared with their high-microglial-binding counterparts. These studies suggest that microglial clearance of extracellular tau-antibody complexes is not the predominant mechanism by which tau seeds are cleared. Therefore, another antibody-mediated mechanism of tau clearance may be at play.

Enter TRIM21, an E3 ubiquitin ligase that resides in the cytoplasm and engages antibody-bound particles and stimulates their clearance through the proteasome. Cell culture studies have demonstrated that the introduction of tau antibodies into cells can induce TRIM21-dependent degradation of the tau-antibody complex (*5*). Does TRIM21 play a role in the clearance of antibody-bound tau seeds in vivo? Mukadam *et al.* used a tau transgenic mouse model that expresses human tau with a pathological frontotemporal lobar dementia–associated mutation. Treatment of organotypic hippocampal slice cultures from these mice with a tau monoclonal antibody, BR134, demonstrated a robust reduction in tau seeding. However, when *Trim21* was deleted, BR134 treatment was incapable of reducing tau seeding. Furthermore, treatment with a polyubiquitin inhibitor also prevented BR134-mediated neutralization of tau seeding, suggesting that TRIM21 is required for tau antibody efficacy and functions by promoting proteasomal clearance of the tau-antibody complex (see the figure).

"…tripartite motif-containing 21 (TRIM21) is required for tau antibody efficacy and functions by promoting proteasomal clearance of the tau-antibody complex…."

Mukadam *et al.* went on to investigate the role of TRIM21 in the efficacy of tau immunotherapy in vivo. Weekly intraperitoneal treatment of tau transgenic mice with the tau-specific monoclonal antibody AP422 resulted in reduced tau pathology and a reduction in seed-competent tau. Conversely, antibody protection was not observed in TRIM21-deficient tau transgenic mice. Behavioral assessment was not conducted, so it is not clear whether mouse cognition was improved upon TRIM21-mediated removal of the tau-antibody complex. Because dementia is a major clinical feature of tauopathies, it is important to investigate this in future studies. Together, these results support a growing body of evidence that microglial activation and phagocytosis of tau-antibody complexes are not required for immunotherapy efficacy. Because microglia have a complex role in tau pathogenesis, with recent studies showing that they are inefficient at clearing tau seeds and can contribute to tau spreading (*6*, *7*), the exclusion of microglia from considerations of future antibody design is welcome news.

Two questions that remain unanswered are how the antibody-tau complex is internalized, and whether the antibody alone can be internalized so that it can then engage intracellular tau. These are important considerations because it has been estimated that only tiny quantities of seed-competent tau exists in the extracellular milieu (*8*). Antibodies designed to target extracellular tau seeds would therefore have to bind with extremely high affinity (in the range of low picomolar to subpicomolar), which can be difficult to achieve (*9*). Perhaps a more promising strategy would be to design antibodies with enhanced neuronal uptake for targeting cytoplasmic tau. Tau antibodies have been shown to be internalized by neurons through engagement with receptors expressed on the surface of neurons or through nonspecific bulk endocytosis (*10*). Furthermore, this process is dependent on antibody charge (*10*). Careful design of the antibody receptor-binding domain and consideration of its charge may facilitate neuronal uptake and subsequent TRIM21-mediated clearance of cytoplasmic tau.

The work by Mukadam *et al.* brings us one step closer to understanding how therapeutic tau antibodies may function in the human brain. Armed with this new information, there is the potential to design second-generation antibodies with enhanced efficacy for the treatment of tauopathies. ■

Florey Institute of Neuroscience and Mental Health The University of Melbourne, Parkville, Melbourne, Australia. Email: rebecca.nisbet@florey.edu.au

REFERENCES AND NOTES

1. A. S. Mukadam *et al.*, *Science* **379**, 1336 (2023).
2. F. Clavaguera *et al.*, *Nat. Cell Biol.* **11**, 909 (2009).
3. R. Bajracharya *et al.*, *Acta Neuropathol. Commun.* **9**, 42 (2021).
4. S. H. Lee *et al.*, *Cell Rep.* **16**, 1690 (2016).
5. W. A. McEwan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 574 (2017).
6. S. C. Hopp *et al.*, *J. Neuroinflammation* **15**, 269 (2018).
7. J. H. Brelstaff *et al.*, *Sci. Adv.* **7**, eabg4980 (2021).
8. C. Sato *et al.*, *Neuron* **98**, 861 (2018).
9. T. E. Golde, *Neurotherapeutics* **19**, 209 (2022).
10. E. E. Congdon *et al.*, *EBioMedicine* **42**, 157 (2019).

ACKNOWLEDGMENTS

R.M.N. acknowledges funding by the National Health and Medical Research Council (GNT2000968) and the Alzheimer's Association (AARG-22-850273).

10.1126/science.adg9800

IMMUNOLOGY

Tuning sentinel immune cells

Neuroimmune interactions in the skin can shape the functions of dendritic cells

By **Barbara U. Schraml[1,2]**

Efficient host defense relies on the ability to mount context-dependent immune responses. Dendritic cells (DCs) sense pathogens and tissue damage and subsequently migrate to lymph nodes to present antigens to naive T cells. Through the production of cytokines, DCs further instruct other immune cells about which type of immune response is needed (*1*). For example, DC-derived interleukin-23 (IL-23) in the skin promotes efficient defense against *Candida albicans* and *Staphylococcus aureus* infections, but it also drives psoriasis-like skin inflammation (*2*, *3*). Nociceptors are somatosensory neurons that innervate barrier organs and detect noxious stimuli, including mechanical injury, reactive chemicals, inflammatory mediators, and pathogens (*4*). Nociceptors relay noxious stimuli to the brain as pain or itching sensation and release neuropeptides, which can influence immune cells (*4*). On page 1315 of this issue, Hanč *et al.* (*5*) report the identification of multiple mechanisms by which nociceptors can regulate DCs in the skin.

The same research group had previously shown that nociceptors potentiate the production of the inflammatory cytokine IL-23 from type 1 and type 2 conventional DCs (cDC1 and cDC2) in imiquimod-induced skin inflammation in mice (*3*). cDC1s are potent activators of $CD8^+$ T cells and promote T helper 1 (T_H1) differentiation, for instance through the production of IL-12. cDC2s efficiently promote T_H2 and T_H17 cell differentiation, the latter in part through the production of IL-23 (*1*). Hanč *et al.* now show that nociceptors promote the production of specific inflammatory cytokines in mouse Fms-related tyrosine kinase 3 ligand (FLT3L)–induced bone marrow DCs. Nociceptors promoted production of IL-12p40, the common subunit for IL-12 and IL-23, and IL-6 but not tumor necrosis fac-

[1]Institute for Cardiovascular Physiology and Pathophysiology, Biomedical Center, Faculty of Medicine, LMU Munich, Planegg-Martinsried, Germany. [2]Walter-Brendel-Centre of Experimental Medicine, University Hospital, LMU Munich, Planegg-Martinsried, Germany. Email: barbara.schraml@bmc.med.lmu.de

ability to reduce tau pathology compared with their high-microglial-binding counterparts. These studies suggest that microglial clearance of extracellular tau-antibody complexes is not the predominant mechanism by which tau seeds are cleared. Therefore, another antibody-mediated mechanism of tau clearance may be at play.

Enter TRIM21, an E3 ubiquitin ligase that resides in the cytoplasm and engages antibody-bound particles and stimulates their clearance through the proteasome. Cell culture studies have demonstrated that the introduction of tau antibodies into cells can induce TRIM21-dependent degradation of the tau-antibody complex (*5*). Does TRIM21 play a role in the clearance of antibody-bound tau seeds in vivo? Mukadam *et al.* used a tau transgenic mouse model that expresses human tau with a pathological frontotemporal lobar dementia–associated mutation. Treatment of organotypic hippocampal slice cultures from these mice with a tau monoclonal antibody, BR134, demonstrated a robust reduction in tau seeding. However, when *Trim21* was deleted, BR134 treatment was incapable of reducing tau seeding. Furthermore, treatment with a polyubiquitin inhibitor also prevented BR134-mediated neutralization of tau seeding, suggesting that TRIM21 is required for tau antibody efficacy and functions by promoting proteasomal clearance of the tau-antibody complex (see the figure).

"…tripartite motif-containing 21 (TRIM21) is required for tau antibody efficacy and functions by promoting proteasomal clearance of the tau-antibody complex…."

Mukadam *et al.* went on to investigate the role of TRIM21 in the efficacy of tau immunotherapy in vivo. Weekly intraperitoneal treatment of tau transgenic mice with the tau-specific monoclonal antibody AP422 resulted in reduced tau pathology and a reduction in seed-competent tau. Conversely, antibody protection was not observed in TRIM21-deficient tau transgenic mice. Behavioral assessment was not conducted, so it is not clear whether mouse cognition was improved upon TRIM21-mediated removal of the tau-antibody complex. Because dementia is a major clinical feature of tauopathies, it is important to investigate this in future studies. Together, these results support a growing body of evidence that microglial activation and phagocytosis of tau-antibody complexes are not required for immunotherapy efficacy. Because microglia have a complex role in tau pathogenesis, with recent studies showing that they are inefficient at clearing tau seeds and can contribute to tau spreading (*6*, *7*), the exclusion of microglia from considerations of future antibody design is welcome news.

Two questions that remain unanswered are how the antibody-tau complex is internalized, and whether the antibody alone can be internalized so that it can then engage intracellular tau. These are important considerations because it has been estimated that only tiny quantities of seed-competent tau exists in the extracellular milieu (*8*). Antibodies designed to target extracellular tau seeds would therefore have to bind with extremely high affinity (in the range of low picomolar to subpicomolar), which can be difficult to achieve (*9*). Perhaps a more promising strategy would be to design antibodies with enhanced neuronal uptake for targeting cytoplasmic tau. Tau antibodies have been shown to be internalized by neurons through engagement with receptors expressed on the surface of neurons or through nonspecific bulk endocytosis (*10*). Furthermore, this process is dependent on antibody charge (*10*). Careful design of the antibody receptor-binding domain and consideration of its charge may facilitate neuronal uptake and subsequent TRIM21-mediated clearance of cytoplasmic tau.

The work by Mukadam *et al.* brings us one step closer to understanding how therapeutic tau antibodies may function in the human brain. Armed with this new information, there is the potential to design second-generation antibodies with enhanced efficacy for the treatment of tauopathies. ■

Florey Institute of Neuroscience and Mental Health The University of Melbourne, Parkville, Melbourne, Australia. Email: rebecca.nisbet@florey.edu.au

REFERENCES AND NOTES

1. A. S. Mukadam *et al.*, *Science* **379**, 1336 (2023).
2. F. Clavaguera *et al.*, *Nat. Cell Biol.* **11**, 909 (2009).
3. R. Bajracharya *et al.*, *Acta Neuropathol. Commun.* **9**, 42 (2021).
4. S. H. Lee *et al.*, *Cell Rep.* **16**, 1690 (2016).
5. W. A. McEwan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 574 (2017).
6. S. C. Hopp *et al.*, *J. Neuroinflammation* **15**, 269 (2018).
7. J. H. Brelstaff *et al.*, *Sci. Adv.* **7**, eabg4980 (2021).
8. C. Sato *et al.*, *Neuron* **98**, 861 (2018).
9. T. E. Golde, *Neurotherapeutics* **19**, 209 (2022).
10. E. E. Congdon *et al.*, *EBioMedicine* **42**, 157 (2019).

ACKNOWLEDGMENTS

R.M.N. acknowledges funding by the National Health and Medical Research Council (GNT2000968) and the Alzheimer's Association (AARG-22-850273).

10.1126/science.adg9800

IMMUNOLOGY

Tuning sentinel immune cells

Neuroimmune interactions in the skin can shape the functions of dendritic cells

By **Barbara U. Schraml**[1,2]

Efficient host defense relies on the ability to mount context-dependent immune responses. Dendritic cells (DCs) sense pathogens and tissue damage and subsequently migrate to lymph nodes to present antigens to naive T cells. Through the production of cytokines, DCs further instruct other immune cells about which type of immune response is needed (*1*). For example, DC-derived interleukin-23 (IL-23) in the skin promotes efficient defense against *Candida albicans* and *Staphylococcus aureus* infections, but it also drives psoriasis-like skin inflammation (*2*, *3*). Nociceptors are somatosensory neurons that innervate barrier organs and detect noxious stimuli, including mechanical injury, reactive chemicals, inflammatory mediators, and pathogens (*4*). Nociceptors relay noxious stimuli to the brain as pain or itching sensation and release neuropeptides, which can influence immune cells (*4*). On page 1315 of this issue, Hanč *et al.* (*5*) report the identification of multiple mechanisms by which nociceptors can regulate DCs in the skin.

The same research group had previously shown that nociceptors potentiate the production of the inflammatory cytokine IL-23 from type 1 and type 2 conventional DCs (cDC1 and cDC2) in imiquimod-induced skin inflammation in mice (*3*). cDC1s are potent activators of $CD8^+$ T cells and promote T helper 1 (T_H1) differentiation, for instance through the production of IL-12. cDC2s efficiently promote T_H2 and T_H17 cell differentiation, the latter in part through the production of IL-23 (*1*). Hanč *et al.* now show that nociceptors promote the production of specific inflammatory cytokines in mouse Fms-related tyrosine kinase 3 ligand (FLT3L)–induced bone marrow DCs. Nociceptors promoted production of IL-12p40, the common subunit for IL-12 and IL-23, and IL-6 but not tumor necrosis fac-

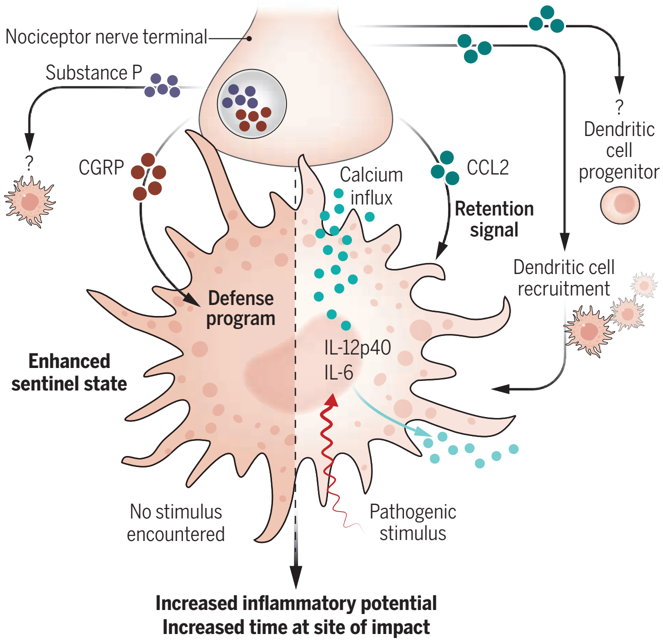
[1]Institute for Cardiovascular Physiology and Pathophysiology, Biomedical Center, Faculty of Medicine, LMU Munich, Planegg-Martinsried, Germany. [2]Walter-Brendel-Centre of Experimental Medicine, University Hospital, LMU Munich, Planegg-Martinsried, Germany. Email: barbara.schraml@bmc.med.lmu.de

tor (TNF), which indicates that nociceptors may modulate specific DC functions. Nociceptors increased cytokine production from DCs in response to various pathogenic stimuli, suggesting that nociceptor-mediated modulation of DCs may be widely applicable to inflammation.

Nociceptors potentiated DC-mediated cytokine production independent of the nociceptor-secreted neuropeptide calcitonin gene–related peptide (CGRP). Hanč *et al.* showed that nociceptors and DCs engage in tight physical interactions in vitro and in vivo. Upon stimulation of nociceptors with capsaicin, an agonist of the transient receptor potential subfamily V member 1 (TRPV1) ion channel that drives pain sensation, they secreted the chemokine C-C motif ligand 2 (CCL2) that attracted DCs, which, in turn, interacted with firing axonal segments that showed membrane depolarization and calcium influx. Overall, nociceptors regulate cytokine production in DCs through electrically mediated membrane depolarization and calcium influx in a contact-dependent manner (see the figure).

Nociceptor modulation of skin dendritic cells

Nociceptors that are activated by noxious stimuli or pathogens secrete neuropeptides such as CGRP, which induce gene programs that enhance DC sentinel functions (defense program) even when DCs have not encountered pathogens. Activated nociceptors attract DCs through CCL2 secretion (retention signal), which facilitates cell-cell interaction and propagation of nociceptor-derived action potentials to DCs. The resulting calcium influx increases the production of cytokines in DCs that have encountered a pathogenic stimulus, and DCs are retained at the site, possibly promoting inflammation.



CCL2, C-C motif ligand 2; CGRP, calcitonin gene–related peptide; DC, dendritic cell; IL, interleukin.

Nociceptors induced gene expression changes in cocultured cDC1 and cDC2 with or without imiquimod stimulation and irrespective of whether nociceptors were activated. These effects could be attributed to CGRP. cDC1 and cDC2 showed up-regulation of gene programs involved in DC functions—such as migration and antigen presentation—suggesting that nociceptors may enhance DC sentinel functions. The proinflammatory cytokine IL-1β is synthesized as a pro-protein (pro–IL-1β). Upon inflammasome activation, pro–IL-1β is proteolytically processed and immediately released. Nociceptor-derived CGRP increased pro–IL-1β expression in DCs, although it did not influence secretion of IL-12p40 or IL-6. Thus, CGRP, in the absence of pathogenic stimuli, renders DCs in a state of enhanced sentinel function, equipping them to exert their functions more rapidly.

Hanč *et al.* demonstrated that electrical activity contributes to the in vivo cross-talk of nociceptors and DCs in imiquimod-induced skin inflammation in mice. Topical treatment with the anesthetics lidocaine and QX-314, which silence nociceptor electrical activity, attenuated inflammation and reduced skin IL-12p40 and IL-1β levels. Topical treatment with capsaicin in the absence of imiquimod induced pro–IL-1β but not IL-12p40 production in DCs, which indicates that nociceptor signaling is insufficient to drive IL-12p40 expression in unstimulated DCs. Loss of CCL2 in nociceptors ameliorated imiquimod-induced skin inflammation and IL-12p40 levels. This effect correlated with fewer DCs in the skin, indicating that nociceptor-derived CCL2 may promote retention of DCs in the skin.

The application of haptens, which are nonimmunogenic small molecules, to skin results in a delayed-type hypersensitivity (DTH) reaction because DCs take up hapten-modified self-antigens and subsequently migrate to and activate T cells in draining lymph nodes. Sensitization by topical hapten application to abdominal skin in conjunction with anesthetics or in mice that lacked CCL2 in nociceptors reduced ear swelling upon rechallenge with the same hapten in the ear. Thus, nociceptors influence the ability of DCs to prime T cells in the lymph nodes, but whether this is a result of transcriptional regulation of genes involved in antigen presentation or a result of CCL2-mediated modulation of DC migration to lymph nodes and prolonging the time that DCs have to acquire antigens remains to be shown.

More work is also needed to determine whether cDC1 and cDC2 are equally regulated by nociceptors because these cell types exhibit distinct immune functions (*1*). Nociceptor-derived substance P promotes migration of CD301b⁺ dermal DCs in a model of skin allergy (*6*) but had little influence on gene expression in the study of Hanč *et al.* This could be because Hanč *et al.* used FLT3L-induced bone marrow DCs in coculture experiments, which do not reflect in vivo DCs. This raises the question of whether nociceptors transcriptionally modulate DCs in vivo. That nociceptor-derived CGRP promoted IL-23 production from CD301b⁺ dermal DCs during *C. albicans* infection (*2*) but not in the study of Hanč *et al.* indicates that nociceptor-mediated regulation of DCs is likely context dependent and influenced by the type of noxious stimulus or pathogen encountered.

The observation that nociceptor-derived CGRP can elicit an enhanced sentinel state in DCs before they encounter pathogens is intriguing and necessitates the determination of whether nociceptor-derived signals affect the development and homeostasis of DCs in the steady state or solely in the context of inflammation or pathogenic insult. In inflammation, nociceptor-derived CCL2 could aid the replenishment of DCs by recruiting DC progenitors from the circulation (*7*). Sensory neurons innervate various tissues, including lymphoid organs, suggesting that neuronal modulation of DCs could be site specific. DC-derived inflammatory cytokines can activate nociceptors to promote itching (*8*). Therefore, DC-nociceptor cross-talk could amplify unwanted sensory or inflammatory responses in settings of chronic pain. Thus, the study of Hanč *et al.* opens new avenues for studying neuroimmune interactions in inflammation. ■

REFERENCES AND NOTES

1. M. Cabeza-Cabrerizo, A. Cardoso, C. M. Minutti, M. Pereira da Costa, C. Reis e Sousa, *Annu. Rev. Immunol.* **39**, 131 (2021).
2. S. W. Kashem *et al.*, *Immunity* **43**, 515 (2015).
3. L. Riol-Blanco *et al.*, *Nature* **510**, 157 (2014).
4. P. Baral, S. Udit, I. M. Chiu, *Nat. Rev. Immunol.* **19**, 433 (2019).
5. P. Hanč *et al.*, *Science* **379**, eabm5658 (2023).
6. C. Perner *et al.*, *Immunity* **53**, 1063 (2020).
7. M. Cabeza-Cabrerizo *et al.*, *Sci. Immunol.* **6**, eabi9331 (2021).
8. J. Xu *et al.*, *Immunity* **53**, 371 (2020).

ACKNOWLEDGMENTS

B.U.S. is supported by grants from the European Research Council (STG-715182) and Deutsche Forschungsgemeinschaft (322359157-FOR2599-A03, 360372040–SFB 1335-P08, and TRR 359-P05-491676693).

10.1126/science.adh2090

GRAPHIC: N. CARY/*SCIENCE*

ASTRONOMY

Streams of cold cosmic fuel for galaxies

An expansive filament of cold gas may connect galaxies and spur star formation

By **Caitlin M. Casey**

Before we understood their scale, origins, or contents, galaxies were dubbed "island universes" for their appearance as isolated beacons of light in an otherwise desolate and dark cosmos. Over the past several decades, observations have revealed that galaxies are connected to one another through a vast network of gas that covers enormous cosmological scales several orders of magnitude larger than individual galaxies (*1*, *2*). This gas—an ethereal tissue on the largest scale—is responsible for forming stars within galaxies. However, the process by which that gas is funneled from a vast cosmic web into galaxies and transformed into stars has proved elusive to observation. On page 1323 of this issue, Emonts *et al.* (*3*) report the detection of a "cold stream" of molecular gas outside of a galaxy, confirming theoretical models that suggested that this route might exist.

It is thought that primordial gas from the Big Bang eventually settled within galaxies into gravitational potential wells (which result from the pull of gravity caused by a body in space) to form stars. However, the path of gas from the primordial cosmic web to the accreting disk of a galaxy has been debated (*4*, *5*). Furthermore, stars can only form from gas under specific conditions—high densities and cold temperatures—and the bulk of gas outside of galaxies exists in the opposite state, diffuse and hot. The process by which hot gas is cooled and deposited into galaxies is thought to take billions of years. During that time, hydrogen gas transforms from an ionized state (caused by irradiation from ambient ultraviolet photons in the cosmos) into a neutral state before transforming into molecular hydrogen that condenses to form stars. Despite the long time scale of this process, galaxies in the young Universe—about 10 billion to 12 billion years in the past—seem to have an overabundance of cold and dense gas, which suggests that there may be some way to circumvent the long cooling process.

Theoretical modeling has suggested that cold gas, which is capable of forming stars on relatively quick time scales, can condense and funnel into galaxies as isolated cold streams (*6*). These streams are thought to be narrow, piercing through a halo of otherwise hot gas surrounding a galaxy. What drives the physics of such a stream is not well understood, but one idea is that cold streams comprise small dwarf galaxies aligned like a string of pearls that fall into a massive galaxy halo (*7*). The existence of dwarf galaxies along a stream could provide a means for gas to cool sufficiently and exist in a molecular state that can form new stars (atomic gas is too warm to enable this). However, it has been exceedingly difficult to confirm the existence of such streams feeding into massive galaxies.

Most gas outside of galaxies is detected by its absorption of light from other astronomical objects. Such a gas can be identified when light from a more distant background source (usually a quasar) is absorbed by gas contained within the halo surrounding a galaxy in the foreground (*8*). However, not all galaxies have bright objects behind them along the line of sight. The probability that this occurs is determined by random chance, and background sources that are sufficiently bright to measure absorption are rare. Furthermore, a cold stream of gas should, in principle, be very narrow (otherwise, it would be destroyed by intergalactic radiation). This further reduces the chance that a background light source will fall along the line of sight. If it does, it may be difficult to infer the stream's shape or orientation from a single sight line.

"These streams are thought to be narrow, piercing through a halo of otherwise hot gas surrounding a galaxy."

Identifying the gas stream in emission spectra by detecting some form of light that traces the extent of the stream is strongly preferable. Yet it is also limited by fundamental observational challenges. Molecular hydrogen lacks a dipole moment and, as a result, is not directly observable at cold temperatures. Instead, astronomers have typically turned to observations of carbon, in either its ionized, atomic, or molecular form (carbon monoxide), to trace the cold gas reservoirs in and around galaxies (*9*). Most such observations are limited to gas inside galaxies because the environment outside of galaxies, replete with ionizing radiation, is thought to be harsh to the existence of cold molecular gas.

Emonts *et al.* used data from the Atacama Large Millimeter/submillimeter Array (a radio interferometer) to discover a cold stream of atomic carbon feeding cold molecular gas into a massive galaxy in the early Universe. The authors collected a spectral and spatial map of a transition of atomic carbon around a distant galaxy, initially intending to study the gas inside the galaxy. They then detected the cold stream outside of the galaxy. The stream stretches across a distance roughly three times larger than the galaxy it is feeding and is sufficiently narrow and collimated to be consistent with cold gas streams seen in cosmological simulations. Having observations that match well to previous predictions from simulations has been long in the making.

The existence of atomic carbon in this cold stream means that it cannot be made up entirely of pristine gas formed after the Big Bang. Carbon, like many elements heavier than helium, is forged in the cores of stars. The presence of atomic carbon on such large scales means that it was ejected from the galaxy in which it formed and is now being recycled, falling back down on a galaxy, which may or may not be the galaxy in which it was originally forged (*10*). The recycling of heavy elements in cold streams hints at another layer of the complexity of how galaxies grow.

The detection of a cold gas stream raises questions about gas feeding into galaxies. It is not clear whether the stream comprises several dwarf galaxies that are not visible to current telescopes. What principles allow cold molecular gas to be suspended intact in a large intergalactic filament? Further observations of cold streams that span a broad range of galaxies are needed to determine the processes that feed gas into galaxies. ■

Department of Astronomy, University of Texas, Austin, TX, USA. Email: cmcasey@utexas.edu

REFERENCES AND NOTES

1. M. McQuinn, *Annu. Rev. Astron. Astrophys.* **54**, 313 (2016).
2. M. E. Putman *et al.*, *Annu. Rev. Astron. Astrophys.* **50**, 491 (2012).
3. B. H. C. Emonts *et al.*, *Science* **379**, 1323 (2023).
4. A. Dekel *et al.*, *Nature* **457**, 451 (2009).
5. R. S. Somerville, R. Davé, *Annu. Rev. Astron. Astrophys.* **53**, 51 (2015).
6. D. Kereš *et al.*, *Mon. Not. R. Astron. Soc.* **363**, 2 (2005).
7. R. Bacon *et al.*, *Astron. Astrophys.* **647**, 107 (2021).
8. J. Tumlinson *et al.*, *Annu. Rev. Astron. Astrophys.* **55**, 389 (2017).
9. C. Carilli, F. Walter, *Annu. Rev. Astron. Astrophys.* **51**, 105 (2013).
10. B. Emonts *et al.*, *Mon. Not. R. Astron. Soc.* **477**, 60 (2018).

10.1126/science.adh1663

RETROSPECTIVE

Garland E. Allen III (1936–2023)

Leading historian of biology and social justice activist

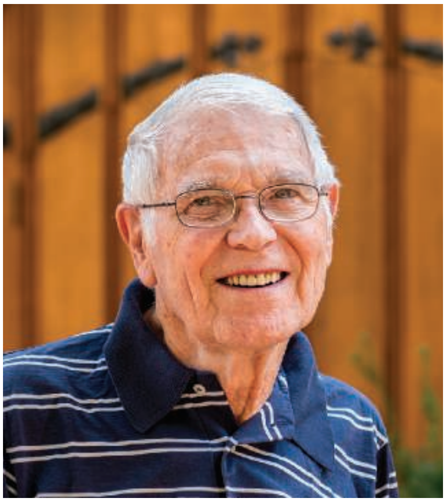
By **Jane Maienschein**

Garland ("Gar") E. Allen III, a leader in the history of genetics and eugenics, died on 10 February at age 86. Gar was at heart an activist, pushing for social justice and promoting inclusion before the concept gained its current urgency. He loved intellectual give-and-take and took pride in the lively philosophical debate generated by the books he authored. A supportive mentor, Gar devoted himself to making vulnerable students feel comfortable in university life.

Born in Louisville, Kentucky, on 13 February 1936, Gar received his bachelor's in English from the University of Louisville in 1957 and a master's in teaching from Harvard University in 1958. After teaching high school for 4 years, he returned to Harvard, where he studied under Everett Mendelsohn and Ernst Mayr as a member of the first generation of historians of biology. After receiving his PhD in 1966, for which he examined the science and life of Nobel laureate Thomas Hunt Morgan, Gar took a position as a professor at Washington University in St. Louis, where he advanced through the ranks and remained until his retirement in 2017.

Gar hated bigotry and exclusionary practices. In the 1960s, he frequently participated in local civil rights demonstrations in St. Louis and, in 1965, traveled to Selma, Alabama, to march with Martin Luther King Jr. He joined the International Committee Against Racism to work toward social and racial justice, and he advocated for social justice in Cuba after an experience harvesting sugarcane there in 1979. Gar used his expertise in the history of genetics to fight against discrimination. Rejecting the concept that one gene could determine a specific trait, he worked to debunk the misconception that homosexuality and race are genetically determined.

In addition to his social justice efforts, Gar improved understanding and inclusion within the science community. His biology textbooks, written with coauthor Jeffrey Baker, and his textbook *Life Science in the Twentieth Century* (1975) reached thousands of students, and his definitive biography, *Thomas Hunt Morgan: The Man and His Science* (1979), inspired generations of historians of science. He brought attention to the social implications of genetics in his George Sarton Memorial Lecture on the problems of genetic determinism and eugenics at the 1998 annual meeting of AAAS (the publisher of *Science*). As president of the International Society for the History, Philosophy, and Social Studies of Biology (ISHPSSB), he helped promote greater diversity and inclusion. In 2017, the History of Science Society awarded him the Sarton Medal, the organization's highest honor for lifetime achievement. This year, he will posthumously receive ISHPSSB's prestigious David L. Hull Prize.



Perhaps some of the greatest beneficiaries of Gar's inclusion efforts were the students at Washington University. Gar provided a sense of belonging, especially for those from underrepresented communities, and he gave generations of students the courage to stand up and speak up, carrying a megaphone and marching for civil rights and anti-war causes with those who were much younger. His vocal political activism put his career advancement at risk, but support within his department allowed him to obtain tenure and promotions.

With his prematurely white hair and slim build, Gar was never intimidating in stature or affect. Graduate students at Washington University and beyond found him to be a wonderful listener and mentor. I first met Gar when he visited my graduate program at Indiana University to discuss his books. An amiable debate about disagreements in interpretation (with Gar characteristically advocating the value of Marxism) began our lifelong friendship. As I learned firsthand, Gar would set trainees at ease, asking about their interests and opinions. The welcoming atmosphere he created allowed him to influence and inspire hundreds of students. Almost every historian of biology has a favorite Gar Allen story, often involving late-night discussions at the pub or at his home.

Gar spent summers at the Marine Biological Laboratory (MBL) in Woods Hole, Massachusetts, where I worked with him for decades. Over the years, he served as MBL Library committee chair and trustee. Starting in 1987, we codirected an annual history of biology course, and we coedited the *Journal of the History of Biology* from 1998 to 2006. When we disagreed, Gar always insisted on a formal debate structure: He would put forth a thesis, and I would counter it with an "antithesis." Gar loved this symmetry, which to him represented a perfect Hegelian "dialectical synthesis," a philosophical approach that led to a positive outcome.

For Gar, Woods Hole meant family, friends, swimming, competitive tennis, tasty meals, sharing ideas, and hospitality. In 1969, he bought a small unheated cottage, and his two daughters, Tania and Carin, could often be seen roaming the town in their early years. In 2022, he continued to shuffle into the Lillie Building's MBL Library each day with his backpack and walker to work on his final book: a history of 20th-century genetics. As he wrote chapters, he corresponded with many of his former colleagues and students, inviting feedback and discussing ideas. Although his health was deteriorating, he did not admit it. Always an optimist, he looked forward to more time and remained insatiably curious. Unfortunately, he did not complete the manuscript, but those of us who were involved in his work will carry forth his ideas.

Gar always gave everybody a chance. His husband—Larry Bennett—and his daughters and grandchildren continue that welcoming, inclusive attitude. Gar made sure that the history of science was not stuffy or overly academic. Rather, he emphasized that historians should always remember the people doing the science and the people affected by science's implications. He was kind and generous, and his friends and former students will miss talking with him. Many a whiskey sour or scotch has been raised in Gar's honor as we continue his work of speaking out against injustice in science, the science community, and the world. ■

10.1126/science.adh5118

Center for Biology and Society, Arizona State University, Tempe, AZ, USA, and Marine Biological Laboratory, Woods Hole, MA, USA. Email: maienschein@asu.edu

PHOTO: SEAN GARCIA/WASHINGTON UNIVERSITY IN ST. LOUIS



PHOTO: WESTEND61/GETTY IMAGES

POLICY FORUM

CHILD DEVELOPMENT

Science for policy to protect children in cyberspace

Cross-disciplinary partnerships are needed to formulate policies that account for developmental vulnerabilities

By Michal Gilad[1,2,3], Diana H. Fishbein[3,4,5], Gideon Nave[6], Nizan Geslevich Packin[1,7,8]

Children (*1*) are not miniature adults. A broad range of scientific studies establish developmental, biological, cognitive, and emotional differences between children and adults that are so substantial that many legal systems recognize their bearing on the rights, liabilities, and protections afforded to children by law. Yet, in some respects, lawmakers are behind the private sector in learning, understanding, and incorporating scientific findings regarding children. The failure to use science-based policy-making and regulations

Many traits of children can be intentionally manipulated to enhance the "stickiness" of apps.

compromises the law's ability to protect children and account for their distinctive attributes, needs, and vulnerabilities. These failures are particularly evident in the protection of children in digital spaces. Despite recognized benefits of some digital technologies for children, these same technologies, when lacking appropriate regulation and monitoring, risk causing substantial harm and can be exploited by entities that prioritize profit over consumers' long-term well-being, especially when vulnerable populations, such as children, are concerned.

Children's use of digital devices is not solely recreational. Devices are an integral part of children's schooling, enrichment activities, community life, and social communication. Therefore, it is unsurprising that studies find continuous drops in the average age of children using such devices, while their daily screen-use steadily increases. For example, empirical surveys conducted before the pandemic showed that more than a third of children in the US began using devices while still in diapers, with the national average of daily screen-time consumption for all children aged 8 and above being between 4 and 9 hours (*2*). These numbers have since increased, as have the types of digital content consumed. Children now engage in digital activities that were previously considered "adults-only," such as financial-technology (FinTech) and wellness apps.

Many tech companies and business platforms have detected, and helped drive, the shifting trends, and the commercial advantages of targeting a younger audience, leveraging the science behind child development. Business entities operating social media, gaming, and FinTech platforms have been able to exploit (intentionally or unwittingly) children's immaturity and vulnerabilities (*3*). Drawing on insights from scientific findings, businesses have identified distinct developmental attributes, including children's impulsivity, underdeveloped decision-making capacities, impressionability, naivety, and susceptibility to peer influence. These traits can be strategically manipulated to enhance the "stickiness" and "engagement" of children with digital products and services (evoking de facto addiction-like compulsion) (*4*) and increase children's consumption and spending. Relying on these findings, these for-profit businesses invested heavily in the design of products, services, and marketing-campaigns that employ research-based tactics that seek to maximize the efficiency and efficacy of their pursuit of children as profitable users. To this end, in 2018 alone, tech-industry players invested over $900 million in science-based marketing strategies targeting children (*5*).

To illustrate the potential risk posed to children in the digital arena, we zoom in with a scientific lens on two commonly used business strategies conducive to negative impacts: gamification, and the incorporation of social features. These examples help us gain a better understanding of how corporate entities capitalize on qualities of the developing child's mind to influence behavior, consumption patterns, and spending.

GAMIFICATION

Gamification, the incorporation of game-like features into nongame apps and services, is a multipurpose strategy that can be used to make products and services more attractive, easy to use, habit-forming, and enjoyable. As such, it is frequently employed in products and services traditionally designed for adults, such as financial services, betting, self-improvement programs, and investment platforms, in an attempt to expand their user base to younger, underage audiences (*6*). To this end, interactive "game-like" elements are used, including scoring, rewards, novelty, peer competition, game rules, quests, and challenges. Such gaming features are familiar and attractive to young users and are effective in gaining their attention during a developmental period of increased distractibility, excitement, sensation seeking, and impulsivity. Such practices can be particularly problematic and harmful when the gamified platforms and services are not developmentally appropriate, involve content and activities that are incompatible with underdeveloped decision-making capacities, and require cognitive perception of long-term implications, comprehension of the subject matter (such as financial decision-making), and parental supervision.

Although there is substantial empirical evidence to support socially beneficial applications of gamification and its ability to enhance motivation, reinforce desired behaviors, and improve learning, cognitive functions and social skills, tech companies too often gamify products to encourage spending and repeat use, rather than to educate. Game designs frequently "play" on children's high risk tolerance and impulsivity to make poor financial decisions, almost in a gambling-like fashion. Moreover, the challenges and rewards featured in games trigger the secretion of adrenaline and dopamine to which the brain is sensitive, particularly during adolescence, reinforcing their thrill-seeking tendencies (*4*). Additionally, the fun play-like nature of the game, coupled with children's limited life experience and immature processing capacities, distracts children from fully comprehending the seriousness and complexity of financial markets and the substantial risk of real monetary loss (*7*).

Also noteworthy is that the traditional gaming industry has begun adopting new business models, which employ manipulative design features meant to increase profit and extend use. Examples include "pay-to-progress" and "pay-to-win" features, which require children to purchase virtual items with real money to advance or win games. These models are further reinforced with peer pressure when strategically placed in competitive multiplayer settings. Research suggests that these engagement strategies may lead to a preoccupation with gaming and gamified activities that resembles substance addiction, allowing for-profit businesses to recruit habitual users and create dependence on their products (*8*). In fact, in recent years, gaming disorder has been added to the International Classification of Diseases (ICD-11) given features in common with addictions and gambling disorder, and on the basis of findings from functional neuroimaging studies showing that similar areas of the brain are activated (*9*).

Existing studies investigating the effects of gamification and gaming on children report a general association between extended engagement with game-like platforms and the negative outcomes. Yet, far less is known about the exact features of the game that cause harm. Research is needed to isolate specific gaming features responsible for harmful effects (e.g., light and sound effects, visual scenes, digital rewards), as well as the potency and mechanisms by which each of these features causes harm. Mediation models and designs such as multiphase optimization strategy (MOST) and Sequential Multiple Assignment Randomized Trial (SMART) have potential to isolate components or features of games that are producing the greatest effect on relevant domains of child development. Such nuanced knowledge can and should inform the design of regulatory policies for such platforms. This information is particularly important as recent attempts to place liability on tech companies for alleged harms and addiction-like behavior associated with gaming and gamification were struck down

[1]The Multidisciplinary Center on Childhood, Public Policy, and Sustainable Society, New York, NY, USA. [2]Salzburg Global Seminar, Washington, DC, USA. [3]National Prevention Science Coalition to Improve Lives, Oakland, CA, USA. [4]Frank Porter Graham Child Development Institute, University of North Carolina-Chapel Hill, Chapel Hill, NC, USA. [5]Pennsylvania State University, State College, PA, USA. [6]Marketing Department, The Wharton School, University of Pennsylvania, Philadelphia, PA, USA. [7]Baruch College, City University of New York, New York, NY, USA. [8]University of Haifa Faculty of Law, Haifa, Israel. Email: nizan.packin@baruch.cuny.edu; michal@mccppss.org

because of their overly broad and nonspecific nature, raising concerns about the potential for interference with fair competition, desirable market incentives, and predictability. Practical evidence-based regulatory directives to tech companies that can withstand current challenges are warranted to effectively balance the need to provide protection without imposing excessive and disruptive liabilities.

SOCIAL FEATURES

Many digital apps, learning programs, and games that are specifically designed for children incorporate social features. These include chat and messaging functions that enable direct interaction with other users (including complete strangers), sharing, and reactions. Such features are designed to appeal to young users' developmental needs for social interaction and belonging, susceptibility to peer influence, and fear of missing out, as well as to incentivize purchases and spending through shaming, unregulated use of social influencers, and peer pressure (*10*). These tactics have a greater influence on children who have a limited capacity to discern advertising biases and can be more easily exploited as consumers. Additionally, these functions often lack sufficient security, adequate privacy protection, and effective parental control mechanisms, leaving the door open for ill-intentioned predators, sexual harassment, bullying, and destructive peer pressure.

Studies suggest that the impact of these unregulated business strategies extends beyond the exploitation of developmental vulnerabilities to modify behavior and elicit profit. Children's exposure to such manipulations, and the protracted use of digital screens they breed, have been found to induce chemical imbalances, alter the structural and functional development of the growing child's brain, and interfere with the ability to meet developmental milestones, often overshadowing the positive effects of new technologies (*11*). Indeed, despite the recognized benefits of social media, gaming, and other digital platforms, the scientific literature is also replete with studies evincing the relationship between childhood exposure to gamification, social media and extended screen time, and diminished social competence, isolation, shifts in values, sleep disruption, aggression, and health and academic difficulties (*12*). In addition, concerning changes in stress biomarkers, including heart rate, blood pressure, and oxygen consumption, have been observed (*13*). These impacts pose a substantial risk of causing harms.

BARRIERS TO TRANSLATION

Such scientific findings are informative and compelling; however, the translation of this body of research into actionable formats for lawmakers requires additional steps and further inquiry to make evident its policy relevance. A majority of these studies are not specifically tailored to answer some of the key questions confronted by policy-makers in this field. Furthermore, the language used in scientific publications is often inaccessible to policy-makers. Moreover, most existing studies establish only association between exposure to technologies and harmful outcomes. More rigorous, carefully controlled studies, using randomized controlled trials or quasi-experimental study designs, are necessary to establish causality and decipher confounding effects (e.g., due to self-selection of individuals into digital platforms), and ground future law and policy-making efforts on a stronger scientific foundation.

Considering the emerging threats posed to children affected by such practices, policy-makers have a fundamental responsibility to apply their parens patriae obligation by providing protections to children, as a designated vulnerable population. Yet, most legislatures and regulatory agencies operating in this domain have largely overlooked scientific evidence in rulemaking efforts. Policy-making is a complex process, influenced by countless competing practical, ideological, procedural, and political factors; thus, there are likely numerous causes behind this failure to act that are difficult to isolate and verify. However, what is observable through analysis of existing policy efforts and proposals is that, absent meaningful collaboration between scientists and law and policy experts, scientific findings remain inaccessible to decision-makers, and research designs are not tailored to address the concrete policy questions at hand. This disconnection may explain the differences between for-profit businesses that are more amenable to drawing upon scientific understanding of children and adolescents as they operate in the digital sphere, in comparison to policy-makers who have been relatively slow to recognize and incorporate the science into policy. Moreover, even the few existing laws that were intended to protect children in the digital era were not designed on the basis of relevant scientific findings, further limiting their efficacy.

For example, agencies such as the Financial Industry Regulatory Authority (FINRA) and the Securities and Exchange Commission (SEC)—which consider addressing harmful

Principles for research and policy

- Academic institutions and grant funders can require, prioritize, and/or reward interdisciplinary research teams and think tanks composed of law and policy scholars and scientists working collaboratively.
- More rigorous, carefully controlled studies, using randomized controlled trials or quasi-experimental study designs, are necessary to establish causality, decipher confounding effects, and isolate specific gaming features responsible for effects.
- Policy evaluation studies will be necessary to look into their efficacy, unintended consequences, and ability to meet changing demands as technologies evolve.
- Grants should include pragmatic requirements, including detailed plans for the translation and dissemination of research findings to policy-makers.
- Interdisciplinary conferences and symposia can facilitate the exchange of real-time research findings, promoting adaptation to the tech industry's fast-changing reality.
- Law and policy experts on research teams can play a key role in translating relevant scientific findings into operational policy instruments and develop persuasive narratives that appeal to and resonate with policy-makers and the general public.
- Beyond external roles as advisers and evaluators, scientists should be integrated into the decision-making process in both the legislative and executive branches of government.
- Federal and state law could require all committees discussing matters related to children in digital spaces to have at least one experienced child development specialist. Such notions have been adopted in other areas of law such as, for instance, financial risk management decision-making.
- Because robust litigation on issues of tech regulation and children is already emerging in courts across the US and abroad, strategies are needed to incorporate and account for the inherent scientific aspects of these cases to increase awareness and understanding in the judicial system.
- Adequate resources and training must be provided to the judiciary to enable judges to properly evaluate the rigor, validity, and weight of scientific evidence pertaining to the issues at stake.

digital "engagement-practices" including the gamification of FinTech services—completely ignore the elevated risk such practices pose to children. Similarly, the Children's Online Privacy Protection Act, formulated to protect children's data privacy, considers children over age 13 as fully consenting adults. However, as recently acknowledged by the US Surgeon General (*14*), setting the bar at age 13 contradicts the scientific consensus that the prefrontal cortex, a brain region that plays an important role in decision-making and impulse control, is not fully developed before age 25. Although neuroscience research suggests age 13 is developmentally ill-prepared for mature decision-making, impulse control, and consequence assessment, further studies are needed to characterize age-based heterogeneity in relevant dimensions of neurodevelopment to help identify the optimal cut-off age in this concrete context of social media use.

In September 2022, California became the first US state to sign a bill addressing the protection of children on digital platforms. Even here, though, the law focuses on privacy and less so on the scientifically proven impacts on children's health and well-being. Also, a proposed provision that sought to hold businesses accountable for designing products that are addictive to children lacked evidentiary foundation to support its validity and importance, and caved to lobbyists' objections (*15*). This effort can be contrasted with the revised UK Online Safety Bill, which is a product of a more elaborate evidence-based process and collaboration with a variety of agencies that relied on scientific work. In comparison to the California Act, the UK Bill, although still imperfect, accounts for the developmental differences between children and adults, and children's heightened vulnerabilities. Thus, it provides enhanced protection to children, including companies' responsibility and criminal liability for the safety of children online. As more and more jurisdictions follow California and the UK's lead and develop different versions of statutory protections for children in cyberspace, policy evaluation studies will be necessary to look into the efficacy of these laws to provide effective protection to children, as well as the unintended consequences of these laws and their ability to meet the changing demands as technologies advance and evolve.

Lastly, compared to the regulated movie and television industries, digital apps' content age ratings have remained unregulated. Surprisingly, existing ratings are based on the subjective discretion of profit-oriented app developers, without any evidence-based guidelines, standards, or unbiased external reviews. In such a reality, age rating is not only inappropriate but also misleading, as ill-informed parents may choose to rely on it when making decisions on developmentally appropriate content, increasing the risk of harmful content exposure.

When the law is uninformed by relevant scientific findings, policy-makers' capacity to proactively identify potential threats, and respond to them in a timely, effective, and developmentally appropriate manner, is inevitably compromised. An evidence-based approach, collaboratively spearheaded by scientists and law and policy experts, results in the ability to preemptively anticipate which types of contents, products, and services are developmentally appropriate for children of different age groups and present possible harms. Hence, products and services found likely to exert unreasonable pressure on children to make impulsive, irresponsible, careless, or unsafe decisions—acting against their own best interest—can be regulated through tax incentives, professional guidelines, penal codes, tort liability, and other means.

MOVING FORWARD

A prerequisite for the success of any evidence-based policy-making effort is the consistent, fruitful partnership between scientists and law and policy experts throughout the entire process—from research conceptualization to policy implementation. The primary objective of scientific research is to build theories, generate knowledge, and develop new technologies, rather than to influence policy and regulatory decisions. Therefore, designing, conducting, and framing dedicated scientific research in a collaborative fashion is critical to generate specific findings with direct implications for policies that protect the health and well-being of children. This will improve upon the current state of affairs, where scientists involved in the policy-making process serve mainly as independent evaluators, or external advisers on scientific advisory boards, providing their input only in later stages of the process. This disjointed dynamic is insufficient to bridge the gap between science and policy, given the vast disciplinary disparities that still exist in language, culture, skill sets, and expertise.

Creating an effective and inherently interdisciplinary translational pipeline that runs throughout the entire evolution of research into operational policy—from the conception of a research question and design to the interpretation of results—is key to increasing the accuracy, relevance, capacity, and likelihood of scientific findings to drive a policy agenda and advance the protection of children. To accomplish this goal, we suggest principles to guide academic institutions and grant funders to influence the direction and utility of research, as well as principles for policy design, adoption, and implementation (see the box). Although success is never guaranteed, such dynamic, continuous collaborations between scientists and law and policy experts could substantially increase the prevalence of effective and child-friendly evidence-based policies. Additionally, even if such policies are not immediately adopted, raising public awareness of these material issues has an instrumental value, nudging consumers to make informed decisions.

Ultimately, interdisciplinary collaboration and lingua franca will promote the translation of scientific findings to practical policies and regulatory instruments and facilitate the formal adoption and effective implementation of resultant policies. Laws informed by scientific evidence would provide more meaningful protection to children using digital platforms, allowing them to safely reap the benefits new technologies provide, while boosting the real-world impact of emerging research. ■

REFERENCES AND NOTES

1. "Children" in this commentary refers to the entire developmental period, from early childhood through to early adulthood (which is generally understood to encompass ages between 18 and 24 years old).
2. See, for example, B. Auxier *et al*., "Children's engagement with digital devices, screen time," *PEW Research Center* (July 2020), https://www.pewresearch.org/internet/2020/07/28/childrens-engagement-with-digital-devices-screen-time; American Academy of Child and Adolescent Psychiatry, "Screen time and children" (February 2020), https://www.aacap.org/AACAP/Families_and_Youth/Facts_for_Families/FFF-Guide/Children-And-Watching-TV-054.aspx#:~:text=Children%20and%20adolescents%20spend%20a,spend%20up%20to%209%20hours.
3. A. Wiltshire, "Behind the addictive psychology and seductive art of loot boxes," *PCGamer* (28 September 2017); http://www.ea.com/en-gb/games/starwars/battlefront/battlefront-2/news/pre-launch-update.
4. A. Braun, "Compulsion loops and dopamine hits: How games are designed to be addictive," *MakeTechEasier* (13 November 2018); https://www.maketecheasier.com/why-games-are-designed-addictive/.
5. J. G. Navarro, "Spending on advertising to children worldwide from 2012 to 2021, by format," *Statista* (January 2023); https://www.statista.com/statistics/750865/kids-advertising-spending-worldwide/.
6. N. Geslevich Packin, *Hastings Law J.* **74**, 349 (2023).
7. A. B. Ortiz de Gortari, J. Gackenbach, *Front. Psychol*. **12**, 585547 (2021).
8. C. F. Watson, *Richmond J. Law Technol*. **27**, 1 (2021).
9. J. B. Saunders *et al*., *J. Behav. Addict*. **6**, 271 (2017).
10. C. Lieber, "Apps for Preschoolers Are Flooded with Manipulative Ads, According to a New Study," *Vox* (30 October 2018); https://www.vox.com/the-goods/2018/10/30/18044678/kids-apps-gaming-manipulativeads-ftc.
11. B. Carpita *et al*., *Life (Basel)* **11**, 775 (2021).
12. M. T. Maza *et al*., *JAMA Pediatr*. **177**, 160 (2023).
13. L. Straker, R. Abbott, *Pediatr. Exerc. Sci*. **19**, 459 (2007).
14. A. Gordon, P. Brown, "Surgeon General says 13 Is 'too early' to join social media," *CNN* (29 January 2023); https://www.cnn.com/2023/01/29/health/surgeon-general-social-media/index.html.
15. Office of Governor Gavin Newsom, Governor Newsom Signs First-in-Nation Bill Protecting Children's Online Data and Privacy, (15 September 2022); https://www.gov.ca.gov/2022/09/15/governor-newsom-signs-first-in-nation-bill-protecting-childrens-online-data-and-privacy/.

10.1126/science.ade9447



Walthamstow Wetlands provides drinking water for London and a haven for breeding and migrating birds.

PHOTO: BLUE SKY IN MY POCKET/GETTY IMAGES

ECOLOGY

Rewilding the city

Urban environments can and should be rich reservoirs of biodiversity

By **Rob Dunn**

Most of us now live in cities where the trunks of trees have been replaced by the gray slabs of buildings and where the calls of wild birds have been overtaken by car horns, engines, and the ever-present murmurations of human voices. In *Urban Jungle*, Ben Wilson reminds readers that cities can be, and have been, otherwise. It is possible to imagine, in his words, "urban jungles," replete with life's riches and benefits.

The word "jungle" derives from the Hindi word "jangal," which descends from the Sanskrit word "jangala." The jangala consisted of savannas managed using fire to favor plant species eaten by pasture animals; they were wild and yet controlled to benefit humans (*1*). A jangal is something different: It is the forested land just beyond human control, a place where species have gone sufficiently feral to prevent human entry.

Humans have often, Wilson argues, thought about the living world in cities as something impenetrable. He quotes Victor Hugo, who wrote, "to observe the city edge is to observe an amphibian. End of trees, beginning of roofs, end of grass, beginning of paving stones, end of ploughed field, beginning of shops…." Human perceptions of this interface have varied though. Wilson shows this variation through many lovely microhistories of the people and decisions that have shaped the biological present, whether in London, New York, Leipzig, or Singapore.

The book begins by considering the tensions between manicured green spaces in cities and wilder patches. From there, Wilson tells the stories of life that presses unbidden through the cracks in cement, the species that thrive despite us. He turns to the value of trees, forests, and rivers before considering what we might eat from the species around us in the city. Of the many animal species that farm—leafcutter ants and termites among them—humans are the only species that grows their food far from where they live. This could be different. In parts of some cities, it already is.



Urban Jungle
Ben Wilson
Doubleday, 2023. 304 pp.

Wilson describes the benefits of nature, including the services that wilder green places in cities can provide to humanity. Trees provide oxygen and filter pollutants, but they also help to buffer cities from sea level rise and extreme heat and protect underground aquifers where wild microbes help to control potential pathogens in drinking water. Riverside vegetation and natural riverbanks prevent flooding. Wild plant species in cities provide food. In some cases, the resources of wild and gardened places in cities can help to conserve charismatic vertebrates, be they flying foxes, parrots, or koalas.

And, of course, the living world of cities can simply be sublime. Today, I spotted a ruby-crowned kinglet in my own city of Raleigh, North Carolina. Little bigger than my thumb, the yellow-tinged bird with a fire red crest wowed me. Urban nature, Wilson reminds us, retains this power.

Wilson's book is full of stories I have not read elsewhere, but it largely focuses on the wild life to which human senses are attuned: plants and vertebrates. This has the potential to lead the reader to imagine that the wild living world is absent in the more industrial areas. It is not. Even in the densest cities, a kind of jungle thrives indoors that is no less wild than a forest. In much of Manhattan—one of the world's greenest urban areas—most of the living mass is made up of human bodies, stored human food, pets, and then, of course, waste. Hundreds of thousands of species live off of this biomass. We might not like them all, but they are part of the urban jungle, part of the corporeal jangal.

The interface between our bodies and the invisible living world in cities matters because some of these species benefit us when present and afflict us when absent. Others can kill us when present. Malaria, dengue, and the virus that causes COVID-19 are parts of wild nature. No one would advocate rewilding our lives with these species. Yet nature is complex. Recent research has shown that bringing more vertebrates into cities can increase the probability of the spread of diseases, whether through increases in the abundance of existing pathogens or through spillover events, the introduction of pathogens to new hosts (*2*).

As we rewild, we need to remember nature's complexities and that the interface between our bodies and the rest of the living world is part of a rapidly unfolding evolutionary drama. Here it is useful to return to the Sanskrit word "jangala," which means not a wild place, but a place in which wild species are managed for their benefit and for ours. Maybe what we need, moving forward, is not an urban jungle, but an urban jangala—cities in which managed life is much wilder than it is now and yet also benefits humanity, supporting, for example, clean drinking water, immune health, and mental health or simply providing the sublimity of a ruby-crowned kinglet searching for insects among the fallen leaves. ■

The reviewer is a professor in the Department of Applied Ecology, North Carolina State University, Raleigh, NC, USA, where he is also the senior vice provost for University Interdisciplinary Programs, and is the coauthor of *Delicious: The Evolution of Flavor and How It Made Us Human* (Princeton Univ. Press, 2021). Email: rrdunn@ncsu.edu

REFERENCES AND NOTES

1. M. R. Dove, *J. Anthropol. Res.* **48**, 231 (1992).
2. R. Gibb *et al.*, *Nature* **584**, 398 (2020).

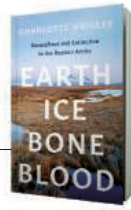
10.1126/science.adg8089

GEOGRAPHY

An ode to Arctic permafrost

A geographer honors a region that defies easy description

Earth, Ice, Bone, Blood
Charlotte Wrigley
University of Minnesota Press, 2023. 256 pp.



By **Lynne Peskoe-Yang**

Despite its name, permafrost is neither permanent nor strictly frozen. Mistranslation explains part of the confusion. The Soviet scientists who helped the Bolsheviks colonize northern Siberia in the early 20th century named the squishy, semicrystalline earth "vechnaya merzlota," a phrase that is almost untranslatable into English but might be better rendered as "perpetually refreezing." (Indigenous Siberians such as the Evenki, Sakhans, and Yakuts have no single word for the kind of earth their territory covers.)



Permafrost advocates film a video pitch that will be used to raise funds for an ecological revitalization project.

What is the best way to describe a substance so ineffable that even its name creates confusion? Geographer Charlotte Wrigley takes on the task of explaining this mysterious terrain to people who likely will never set foot on it in her bracing debut, *Earth, Ice, Bone, Blood*. Part of the book's overall message is that Russia's Arctic permafrost region, like any ecosystem, becomes more complicated, not less, the longer one considers it.

Permafrost, in Wrigley's view, takes the form of "scientific knowledge, of hypothesis, of apocalypse, of story and myth… as preserver of history and destroyer of futures…as a delineation on a map or a reading on an instrument, the feeding ground of a Kalmykian cow and the resting ground of a long-dead mammoth god." She doggedly pulls apart the basic vocabulary of geology, geography, and evolutionary science, arguing convincingly that using words such as "ecosystem" and "extinction" uncritically will only reinforce their generalizations. The environmental science buzzword "Anthropocene," for example, which defines our current age in terms of the influence of human beings, is insufficient, she argues, because it "defines the 'Anthropos' as a homogeneous category." Likewise, a word such as "apocalypse" falters because it implies a universal future experience. Many groups have already lived through catastrophic displacement and genocide, reminds Wrigley.

Most of the people who appear in the pages of *Earth, Ice, Bone, Blood* are scientists like Wrigley, whose subtle knowledge of the Arctic permafrost is perpetually challenged by the pressures of communicating their work. In one passage, a group of researchers working on a permafrost conservation program built on decades of research in geoengineering and paleontology attempt to condense their project into a one-minute fundraising video. The work involves relocating some very confused Danish bison to Siberia in the hopes that their hooves will fill the permafrost-preserving ecological niche once occupied by mammoths. But the project needs funding, and apocalypses sell. Eventually, the scientists title the fundraiser "Bison to Save the World." The semantic break between what researchers and locals experience on the permafrost and the way they describe it to the uninitiated is an example of the book's genuinely useful concept of "discontinuity"—the notion that something is impossible to summarize.

Discontinuity colors the work of researchers who study the permafrost. Wrigley describes dinnertime at the permafrost research center, when scientists of all stripes assemble after a long day of trudging through mud and snow to reach the specific segments of permafrost that serve as their laboratories. The conversation is "centered on permafrost in a rather patchwork way made up of divergent knowledges. We speak anecdotally, share rumors and gossip, and extrapolate how climate change will affect permafrost next year, in ten years, in fifty."

Wrigley's prose is appropriately discontinuous. On one page, permafrost "slumps, creeps, karsts, shears, heaves; it forms polygons, contains ice wedges, produces lakes and rich *alaas* ecosystems…[it can] boil, blister, burst, churn, crack, fissure, pry, shatter, shift, spit, and suck." Elsewhere, phrases such as "liminal material state," "intersecting multiplicities," and "dynamic and discontinuous processes" seem designed to test a reader's commitment.

Acknowledging discontinuity makes it possible to study something without having to claim exclusive or complete knowledge of the subject. Environmental science is sorely in need of this approach. When confronting something as huge and intricate as permafrost or the climate crisis, nuance and humility can offer an escape route from the plague of generalization. While by no means an easy read nor an antidote to the climate blues, *Earth, Ice, Bone, Blood* rewards its readers with its sensory experience and its philosophical meditations, arming them with new questions with which to challenge their own slow-churning surroundings. ■

10.1126/science.adf6999

The reviewer is a freelance science journalist based in Massachusetts. Email: lynnepeskoeyang@gmail.com

PHOTO: COURTESY OF THE UNIVERSITY OF MINNESOTA PRESS

LETTERS

Edited by **Jennifer Sills**

Retraction

We have recently been made aware of an inconsistency in the data presented in the Report "A shorter ^{146}Sm half-life measured and implications for ^{146}Sm-^{142}Nd chronology in the Solar System" (*1*). After further examination of the records, we attribute the discrepancy to a flaw in the chemical processing of the Sm samples analyzed.

In the Report, we measured the ^{146}Sm half-life by determining the double ratio of atom number to alpha activity for ^{146}Sm to ^{147}Sm and using the known half-life of naturally occurring ^{147}Sm. The ^{146}Sm and ^{147}Sm activity ratios were measured in thin sources made from ^{147}Sm-enriched material irradiated to produce ^{146}Sm. The ^{146}Sm and ^{147}Sm atom ratios were measured in samples of larger mass obtained from dilution of the original thin sources with ^{nat}Sm.

The dilution ratio is required to derive the ^{146}Sm and ^{147}Sm atom ratio in the original Sm sources from that measured in the diluted samples. We now recognize a discrepancy between the dilution ratios used in the Report, listed in table S1 of the supporting online material (SOM), and values that can be independently extracted based on the ^{147}Sm alpha activities (also listed in table S1), the ^{147}Sm half-life, and the dilution procedure detailed in SOM.

The dilution ratios listed in table S1 were obtained from a measurement of the ^{147}Sm content of the original sources done by inductively coupled plasma mass spectrometry (ICP-MS). We attribute the observed discrepancy to an unrecognized systematic error in the ICP-MS measurement. A reliable reanalysis of the 2012 data or correction of the results is not possible at this time. We therefore retract the Report. All authors agree to the Retraction except for D. J. Henderson and T. Nakanishi, who could not be reached.

N. Kinoshita[1], M. Paul[2*], Y. Kashiv[3], P. Collon[3], C. M. Deibel[4], B. DiGiovine[5], J. P. Greene[6], C. L. Jiang[6], S. T. Marley[4], R. C. Pardo[6], K. E. Rehm[6], D. Robertson[3], R. Scott[6], C. Schmitt[3], X. D. Tang[7], R. Vondrasek[6], A. Yokoyama[8]
[1]Shimizu Corporation Institute of Technology, Tokyo, Japan. [2]Hebrew University of Jerusalem, Jerusalem, Israel. [3]University of Notre Dame, Notre Dame, IN, USA. [4]Louisiana State University, Baton Rouge, LA, USA. [5]Los Alamos National Laboratory, Los Alamos, NM, USA. [6]Argonne National Laboratory, Argonne, IL, USA. [7]Institute of Modern Physics, Lanzhou, China. [8]Kanazawa University, Kakumamachi, Kanazawa, Ishikawa, Japan.
*Corresponding author. Email:paul@vms.huji.ac.il

REFERENCES AND NOTES

1. N. Kinoshita *et al.*, *Science* **335**, 1614 (2012).

10.1126/science.adh7739

PHOTO: RGB VENTURES/SUPERSTOCK/ALAMY STOCK PHOTO



Insufficient data privacy policies could undermine public trust in newborn screening programs.

Protect newborn screening programs

Nearly every baby born in the United States undergoes a mandatory heel stick to collect blood used for screening tests to identify rare but serious health conditions. Early detection offers the opportunity to prevent death or disability through prompt intervention and treatment—a condition is identified in about 12,500 infants each year (*1*). However, as a recent lawsuit in New Jersey exemplifies (*2*), law enforcement access to residual newborn blood spots and related gaps in privacy protections are poised to undermine such programs, which the Centers for Disease Control and Prevention has called one of the top 10 public health achievements of the 21st century (*3*). Given the substantial benefits of newborn screening programs, states must earn public trust by strengthening data privacy policies.

Last year, the New Jersey Office of the Public Defender alleged that law enforcement officials had subpoenaed the state's newborn screening program to provide the blood spot of a child whose father was accused of committing a sexual assault (*4*). According to reports, law enforcement did not meet the standard of probable cause to obtain a warrant to test the suspect directly before accessing the blood sample (*2*). Later reporting revealed that four New Jersey police departments have used newborn blood spots in five criminal cases (*5*). Other states have had similar requests from investigative agencies. In 2020, multiple warrants and court orders seeking law enforcement access to newborn blood spots were made in California (*6*). In Texas and Minnesota, lawsuits have led to the destruction of biospecimens and the loss of critical public health data in response to concerns about parental consent and privacy (*7*).

Overly permissive biospecimen- and data-sharing policies jeopardize the benefits of newborn screening. State legislatures and the public already misunderstand and underappreciate screening programs (*8*, *9*), and privacy gaps and law enforcement access further undermine public trust. Police access to newborn blood spots will likely unfairly and disproportionately surveil Black men, as has been the case for other methods of DNA sample collection by law enforcement [e.g., (*10*)].

To protect the future of newborn screening and the vital research it supports, states should work to amend their newborn screening policies, both to limit law enforcement access and to provide additional clarity for parents about how their child's tissue and data will be stored and shared (*11*). Parental opt-in consent should be required for any storage and use of identifiable data outside of its original purpose. Any potential risk of law enforcement access should be clearly communicated. At a time when many public health programs are under serious threat, it is imperative to ensure that lifesaving newborn screening programs are worthy of the public's trust.

Kellie Owens[1*], Carolyn Chapman[1,2], Arthur Caplan[1]
[1]Division of Medical Ethics, Department of Population Health, New York University Grossman

School of Medicine, New York, NY, USA. [2]Center for Human Genetics and Genomics, New York University Grossman School of Medicine, New York, NY, USA.
*Corresponding author.
Email: kellie.owens@nyulangone.org

REFERENCES AND NOTES

1. Centers for Disease Control and Prevention, *MMWR Morbid. Mortal. Wkly Rep.* **61**, 390 (2012).
2. N. Biryukov, "Newborn screening program used to aid criminal investigation, public defender says," *New Jersey Monitor* (2022).
3. R. Koppaka, *MMWR Morbid. Mortal. Wkly Rep.* **60**, 619 (2011).
4. NJ Office of the Public Defender *et al*. vs. Department of Health *et al*. (2022); https://www.documentcloud.org/documents/22084922-nj-office-of-the-public-defender-et-al-vs-department-of-health-et-al.
5. D. Difilippo, "Judge orders state to release information about police use of baby blood spots," *New Jersey Monitor* (2023).
6. J. Watts, "CBS13 investigates: CA still storing newborn DNA without consent, Golden State Killer case raising new concerns," *CBS Sacramento* (2020).
7. N. Ram, *Texas Law Rev.* **100**, 1253 (2021).
8. S. M. Andrews, K. A. Porter, D. B. Bailey Jr., H. L. Peay, *BMC Pediatr.* **22**, 90 (2022).
9. S. E. McCandless, E. J. Wright, *Birth Defects Res.* **112**, 350 (2020).
10. J. Ransom, A. Southall, "'Race-biased dragnet:' DNA from 360 Black men was collected to solve Vetrano murder, defense lawyers say," *The New York Times* (2019).
11. Institute of Medicine Roundtable on Translating Genomic-Based Research for Health, "Challenges and opportunities in using residual newborn screening samples for translational research: Workshop summary" (National Academies Press, 2010), ch. 4; https://www.ncbi.nlm.nih.gov/books/NBK52738/.

10.1126/science.adh2746

The science meritocracy myth devalues women

As Women's History Month comes to a close, we shouldn't overlook the women who contribute so much to science and yet continue to be devalued, ignored, and discouraged from pursuing an academic career. Women researchers are cited less often than men (*1*) and are substantially less likely than men to be credited with authorship within research teams (*2*). A variety of factors contribute to gender disparity in science, including harassment (*3*), conscious and unconscious gender bias (*4*), and motherhood (*5*). One often-overlooked obstacle is the myth of meritocracy (*6*), a set of assumptions that obscures the challenges faced by women and other underrepresented groups.

The idea that academic success is solely based on individual merit and hard work is deeply ingrained in scientific and academic systems. The meritocracy myth presumes that those who rise to tenured and leadership positions are more capable or prioritize research more highly than those who do not. However, systemic barriers, including biases and discrimination, often prevent women from advancing in their careers regardless of their skills or commitment (*7*).

To dismantle the meritocracy myth, institutions must acknowledge that hiring and promotion decisions are subject to explicit and implicit biases as well as expectations that discriminate against certain groups (*8*). For example, individuals may unconsciously favor candidates who fit stereotypical ideas of what a successful academic should look like. In addition, expectations that scientists will work long hours are more difficult to meet for those who bear the brunt of caregiving responsibilities at home—most often women (*9*).

To increase women's participation in science, academia must work toward equitable career advancement opportunities. Metrics of success should include mentorship, community engagement, and social impact (*10*) as well as traditional measures such as number of publications and citations. Measuring success should also involve a holistic approach that considers the individual's experiences and recognizes the barriers and challenges they may have faced. A true meritocracy would value individuals for their skills and work ethic without penalizing them for their background, identity, or personal obligations. We can't let misguided assumptions make the pursuit of science untenable for the next Marie Curie, Rosalind Franklin, or Gladys West.

Fernanda Staniscuaski
Federal University of Rio Grande do Sul, Brazil, and The Parent in Science Movement, Porto Alegre - RS, 91501-970, Brazil.
Email: fernanda.staniscuaski@ufrgs.br

REFERENCES AND NOTES

1. C. Lopez Lloreda, *Science* 10.1126/science.caredit.adf306313 (2022).
2. M. B. Ross *et al*., *Nature* **608**, 135 (2022).
3. National Academies of Sciences, Engineering, and Medicine, "Sexual harassment of women: Climate, culture, and consequences in academic sciences, engineering, and medicine" (The National Academies Press, 2018).
4. C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham, J. Handelsman, *Proc. Natl. Acad. Sci. U.S.A*. **109**, 16474 (2012).
5. P. B. M. Carpes, F. Staniscuaski, L. Oliveira, R. C. Soletti, *Epidemiol. Serv. Saúde* **31**, e2022354 (2022).
6. M. Blair-Loy, E. A. Cech, *Misconceiving Merit: Paradoxes of Excellence and Devotion in Academic Science and Engineering* (University of Chicago Press, 2022).
7. R. L. Roper, *Microbiol. Mol. Biol. Rev*. **83**, e00018 (2019).
8. I. Régner, C. Thinus-Blanc, A. Netter, T. Schmader, P. Huguet, *Nat. Hum. Behav*. **3**, 1171 (2019).
9. S. Jolly *et al*., *Ann. Intern. Med*. **160**, 344 (2014).
10. S. W. Davies *et al*., *PLOS Biol*. **19**, e3001282 (2021).

10.1126/science.adh3071

Institutions' role in trainee success

In their Editorial "Students and postdocs deserve more" (10 February, p. 519), S. M. Malcom and S. Parikh describe elements of academia's culture and governance that have stirred justified distress and protest among our trainees. I disagree, however, with their conclusion that funding agencies alone can address the problems. Funders indeed have important capacities, including increasing stipends and benefits for students and postdocs, but academic research institutions also hold both power and responsibility to effect change.

To improve the academic environment for students and postdocs, institutions must look inward. Universities should reassess and overhaul policies, practices, and incentives that drive faculty behavior. Criteria for promotion and tenure should be established that reward contributions to the next generation of scientists. For example, such decisions should value teaching, mentoring, and successful trainee career trajectories. Deans and chairs should celebrate rather than discourage risky, bold research projects that inspire and motivate trainees. Collaborative, open, transdisciplinary investigations should be acknowledged for capitalizing on the scope, impact, and excitement of team pursuits. Administrations should recognize faculty who organize career exploration programs that effectively equip students to select with confidence among the myriad occupations that benefit from the depth and skills of PhD-trained scientists.

As Malcom and Parikh explain, institutional reflexes are driven by an appetite for grant dollars. Deans and chairs urge their faculty to focus on research output over mentorship because publications bring in grants, and they favor conservative, pedestrian projects that align with the risk aversion of funders. The cruel irony here is that students choose to do science because they aspire to achieve something bold and amazing, only to be counseled by their advisers, chairs, and deans (all of whom shared the same aspirations when they started training) to tamp down their goals.

Yes, funders are accountable here. But there is much that institutional leadership, independent of funders, can do to incentivize the kind of mentorship, creativity, and opportunities that young scientists need and deserve. Crafting academic research environments with trainees at the center will in turn strengthen our science and technology enterprise and its essential role in society.

Keith R. Yamamoto
Science Policy & Strategy, University of California, San Francisco, San Francisco, CA, USA. Email: yamamoto@ucsf.edu

COMPETING INTERESTS

K.R.Y. is president of AAAS, the publisher of *Science*.

10.1126/science.adh3723

RESEARCH

IN *SCIENCE* JOURNALS

Edited by
Michael Funk



Analysis of theropod skulls, such as the *T. rex* skeleton pictured here, suggests that these dinosaurs had lips that covered their teeth.

PALEONTOLOGY

Not a toothy grin

Theropod dinosaurs such as the iconic *Tyrannosaurus rex* have long been portrayed with their teeth fully visible, similar to extant crocodilians. This pattern of portrayal largely had to do with relatedness between dinosaurs and crocodilians and the relationship between tooth and jaw size. Cullen *et al*. tested hypothesized facial reconstructions in this group using histological analysis of tooth wear patterns and quantitative relationships between skull length and tooth size in both extinct and extant reptiles. Contrary to depictions that have dominated for more than a century, they found that theropods, including *T. rex*, had lips that covered their teeth, leaving them looking more like modern Komodo dragons than crocodiles. —SNV

Science, abo7877, this issue p. 1348

NEUROSCIENCE

A metabotropic glycine receptor

Until now, ion channels were the only receptors known to mediate the inhibitory effects of glycine. However, glycine can also exert modulatory metabotropic effects through as-yet unclear mechanisms. Laboute *et al*. discovered that the orphan receptor GPR158 acts as a metabotropic glycine receptor. In analogy to the well-known metabotropic glutamate receptors (mGluRs), they named it mGlyR. This new receptor is a member of the G protein–coupled receptor family, which signals by altering the concentration of the central second messenger cyclic adenosine monophosphate. Glycine is directly recognized by a ligand-binding Cache domain present in mGlyR and regulates the activity of cortical neurons. This work introduces an additional neuromodulatory component that will give further insights into synaptic transmission. —PRS

Science, add7150, this issue p. 1352

APPLIED PHYSICS

All fiber fluidics

Fibers are the basis of ropes and fabrics. With the development of advanced fibers that can act as sensors and actuators or store or harvest energy, it is possible to make fabrics with advanced functionalities. Smith *et al*. show that it is also possible to design fibers to pump fluids, a task otherwise left to external, bulky devices. Helical electrodes are first embedded into the walls of a tubular fiber. Then, the injection of charge using a direct current field ionizes a dielectric liquid, and the motion of ions toward the positive electrode drives fluid flow. —MSL

Science, ade8654, this issue p. 1327



Helical electrodes pump cool fluid (blue) within tubes in a wearable thermoregulatory device.

FUNCTION PREDICTION

A crystal ball for enzyme activities

With rapidly growing genomic and metagenomic databases, we have vastly more sequence data than functional data for enzymes. Accurate functional annotation from sparse experimental evidence is therefore crucial for analysis and applications when working

PHOTOS: ANTIC ANDREJ/EYEEM/GETTY IMAGES; M. SMITH ET AL.

from sequence data. Hoping to circumvent the limitations of current approaches, Yu *et al*. developed a machine learning model based on contrastive learning that performs particularly well at discerning enzyme function. In addition to comparing the performance of the method with existing tools, the authors experimentally validated predicted functions of 36 enzymes that form carbon–halogen bonds. They found excellent prediction accuracy and the ability to distinguish between similar activities. —MAF

Science, adf2465, this issue p. 1358

ORGANIC CHEMISTRY

Elevating Wacker to a higher state

Wacker oxidation uses palladium catalysis to produce aldehydes or ketones from olefins and water. The generally accepted mechanism relies on a hydride migration, which is not possible in geminally disubstituted olefins, which lack the requisite hydrogen. Feng *et al*. report a method to promote carbon migration in this class of substrates by oxidizing a catalyst intermediate to the Pd^{IV} state. The resulting pathway converts a wide variety of alkyl and aryl olefins into ketones. —JSY

Science, adg3182, this issue p. 1363

HUMAN GENOMICS

Dominance in the Biobank

According to classical genetics, traits are usually divided into dominant and recessive, with all-or-none phenotypes. Such binary traits are often observed in plant and animal breeding, but they are rarely seen in human studies outside of monogenic disorders with Mendelian inheritance patterns. Instead, many traits included in human genomic studies exhibit additive effects based on gene dosage. Palmer *et al*. examined data from the UK Biobank and identified gene loci showing dominance effects in which the phenotypes are not explained by a purely additive pattern. The authors identified 175 such loci and provided computational tools for others to perform similar analyses of dominance. They also cautioned that reliable detection of such loci requires larger sample sizes and statistical power than for purely additive effects. —YN

Science, abn8455, this issue p. 1341

NEUROSCIENCE

A nuclear exit for learning and memory

Nuclear accumulation of the second messenger cyclic adenosine monophosphate (cAMP) elicits changes in gene expression in hippocampal neurons that underlie learning and memory. Martinez *et al.* found that stimulation of the β_2-adrenergic receptor induced the nuclear export of the cAMP-degrading phosphodiesterase PDE4D5 (see the Focus by Gurevich and Gurevich). This effect required phosphorylation of the β_2-adrenergic receptor by G protein–coupled receptor kinases (GRKs). Mice expressing a mutant receptor lacking the GRK phosphorylation sites had memory deficits that were attenuated by a PDE4 inhibitor. —WW

Sci. Signal. **16**, eade3380, adg9504 (2023).

SOIL RESPIRATION

The exhalation of drylands

Semi-arid regions cause much of the interannual variability in global atmospheric carbon dioxide concentration because of seasonal fluctuations in water availability. Metz *et al*. found that recurrent carbon dioxide pulses occur at the end of the dry season over Australia, an effect that they attribute to enhanced soil respiration preceding photosynthetic uptake in semi-arid regions. The magnitudes of these pulses are two to three times larger than previous studies have inferred based on in situ measurements. —HJS

Science, add7833, this issue p. 1332

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**



PHOTO: ANDREW SYRED/SCIENCE SOURCE

PLANT SCIENCE

Splitting hairs

The leaves and shoots of tomato plants produce seven types of hairs, which can be grouped into digitate (finger-like) trichomes and peltate (shield-shaped) trichomes bearing multicellular glands. Wu *et al.* found that the concentration of a transcription factor called Woolly regulates the differentiation of trichomes. High levels of Woolly promote digitate trichome formation by up-regulating and forming a complex with other transcription factors called WOX and MYB. This module suppresses *LEAFLESS* expression, which would otherwise trigger peltate trichome development. In peltate trichomes, a negative feedback loop maintains low levels of Woolly to allow LEAFLESS activity. This work highlights how bistability arises from a concentration-dependent mechanism to influence cell fate specification. —MRS

Dev. Cell (2023) 10.1016/j.devcel.2023.01.009.

Colored scanning electron micrograph of trichomes on a tomato stem, which develop into multicellular glandular (brown) or digitate forms

PESTICIDES

Falling into trouble

The global spread of pests such as the fall armyworm is pressuring smallholder farmers in less developed countries to resort to unsafe pesticide use. Recognizing the lack of training facilities for maize-growing households in Rwanda and Uganda, Tambo *et al.* tested the effectiveness of mass media campaigns for educating farmers about the dangers of pesticides. They tested the relative value of radio broadcasts, text messaging, video screening, and plant health rallies on pest management. Questionnaires revealed that radio broadcasts were the most popular source of information, although, concerningly, the survey revealed

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

IMMUNOLOGY

Connecting autoimmunity and obesity

Evidence suggests that obese individuals have a higher risk of developing various autoimmune diseases such as multiple sclerosis and type 1 diabetes. In a Perspective, Matarese proposes a possible unifying mechanism that could explain this connection. He suggests that metabolic overload can suppress immunoregulation and at the same time induce inflammatory responses, thereby increasing the risk of autoimmunity. Given this potential immunometabolic feedback, it is possible that altering diet, caloric intake, or drugs that mimic these conditions could modulate autoimmunity. —GKA

Science, ade0113, this issue p. 1298

MOIRE PHOTONICS

Patterning optoelectronics

Moiré patterns develop when two or more periodic lattices are overlaid over each other. These long-range patterns can also be seen in two-dimensional materials and give rise to rich physics within these quantum materials systems. Du *et al*. reviewed recent developments in moiré-patterned two-dimensional materials with a focus on the photonic and optoelectronic aspects and how the properties can be controlled by engineering the stacking order and twist angle between the layers. The wealth of properties seen indicates that these systems may be a powerful platform to develop an exotic optoelectronic device technology. —ISO

Science, adg0014, this issue p. 1313

PLANT SCIENCE

Hormonal control of root elongation

Hormones known as brassinosteroids affect many aspects of plant growth and development. Nolan *et al*. applied single-cell RNA sequencing to the roots of the small mustard plant *Arabidopsis* to study how brassinosteroids control the developmental shift that cells undergo as they cease proliferation and begin to elongate. The gene network involved drove altered expression of genes involved in cell wall biogenesis and organization. Two particular transcription factors were identified as regulators of cell elongation. —PJH

Science, adf4721, this issue p. 1314

NEUROIMMUNOLOGY

Nociceptors and dendritic cells tango

Nociceptors are afferent neurons that transmit pain and itch sensations in response to noxious stimuli. Although they can influence dendritic cells (DCs), the specifics of nociceptor–DC cross-talk have been unclear. Hanč *et al*. found that nociceptors can control DC functions in at least three ways that are context dependent (see the Perspective by Schraml). Nociceptors can attract DCs to tissues and regulate how long they stay there through the chemokine CCL21. When triggered, nociceptors can release the neuropeptide CGRP, which induces a "poised" DC gene program. Finally, nociceptors can trigger calcium ion flux and membrane depolarization in DCs, amplifying DC inflammatory responses through direct electrical coupling. This system allows the immune system to integrate pain with inflammatory and infectious signals in an exquisitely fine-tuned manner. —STS

Science, abm5658, this issue p. 1315; see also adh2090, p. 1301

PHYLOGEOGRAPHY

Making a horse culture

Horses evolved in North America and dispersed to Eurasia across the Bering Land Bridge. They continued to evolve and were domesticated in Eurasia, but, as far as we know, they became extinct in North America by the late Pleistocene and were then reintroduced by European colonizers. Taylor *et al*. looked at the genetics of horses across the Old and New Worlds and studied archaeological samples. They found no evidence for direct Pleistocene ancestry of North American horses, but they did find that horses of European descent had been integrated into indigenous cultures across western North America long before the arrival of Europeans in that region. —SNV

Science, adc9691, this issue p. 1316

GALAXIES

An intergalactic gas stream at cosmic noon

Galaxies grow by accreting material, either in mergers with other galaxies or from gas in the intergalactic medium. Emonts *et al*. used submillimeter observations to map the atomic carbon gas around a massive galaxy at redshift 3.8, early in a period known as cosmic noon, when galaxies were rapidly assembling (see the Perspective by Casey). They identified a stream of cold gas several times the size of the galaxy extending away from it into the intergalactic medium. The mass of gas in the stream would be sufficient to maintain star formation in the galaxy for hundreds of millions of years if it is accreted. —KTS

Science, abh2150, this issue p. 1323; see also adh1663, p. 1303

TAU IMMUNOTHERAPY

Tau and TRIM21

Aggregation of the protein tau inside neurons contributes to common neurodegenerative conditions, including Alzheimer's disease. Mukadam *et al*. demonstrate that mouse models of tau pathology can be treated using anti-tau antibodies by engaging the cytosolic antibody receptor TRIM21 (see the Perspective by Nisbet). Tau assemblies can transit into cells with antibodies attached, thereby bringing TRIM21 and antibody into contact. As an atypical receptor, TRIM21 possesses E3 ubiquitin ligase activity and can mediate the inactivation of these tau assemblies that would otherwise act as seeds for further aggregation. The discovery of this pathway could enable optimization of future immunotherapies. —SMH

Science, abn1366, this issue p. 1336; see also adg9800, p. 1300

VIRAL INFECTIONS

Yellow fever virus treatment

Yellow fever virus (YFV) is once again becoming a threat to global health because of changes in the environment. A live attenuated vaccine (YFV-17D) has been used for decades, but immunization coverage diminished in recent years because of concerns about adverse reactions, which has left many unvaccinated people in endemic areas vulnerable to infection. Ricciardi *et al*. characterized monoclonal antibodies (mAbs) that can neutralize YFV and prevent severe disease in animal models. They screened 37 YFV-specific mAbs isolated from vaccinated humans, of which two showed potent neutralization activity against a range of YFV isolates. Administration of a single dose of either mAb to YFV-infected hamsters or nonhuman primates protected against severe disease and death. These findings suggest that YFV mAbs may be a potential therapeutic treatment that could improve outcomes during outbreaks. —CNF

Sci. Transl. Med. **15**, eade5795 (2023).

ANTIBODIES

Immunoglobulin indel insights

Insertions and deletions (indels) are genomic DNA alterations that are critical to the emergence of a long complementarity determining region 3 (CDR3), which is a key step during immunoglobulin diversification. Indels have been notoriously difficult to detect, thus making it difficult to define mechanisms involved in this process. Hao *et al*. developed *passenger-immunoglobulin* mouse models to carry out ultradeep profiling of indels. They observed that activation-induced cytidine deaminase–dependent single-base-pair indel events were the most common and involved base excision repair. By contrast, longer in-frame indels, such as those associated with CDR3 lengthening, were rare and required other DNA-processing enzymes. The authors were also able to increase indel frequencies through ectopic expression of a DNA exonuclease or changes in the balance of DNA polymerases, thus suggesting a method for generating antibodies with enhanced neutralization. —CNF

Sci. Immunol. **8**, eade1167 (2023).

from sequence data. Hoping to circumvent the limitations of current approaches, Yu *et al*. developed a machine learning model based on contrastive learning that performs particularly well at discerning enzyme function. In addition to comparing the performance of the method with existing tools, the authors experimentally validated predicted functions of 36 enzymes that form carbon–halogen bonds. They found excellent prediction accuracy and the ability to distinguish between similar activities. —MAF

Science, adf2465, this issue p. 1358

ORGANIC CHEMISTRY

Elevating Wacker to a higher state

Wacker oxidation uses palladium catalysis to produce aldehydes or ketones from olefins and water. The generally accepted mechanism relies on a hydride migration, which is not possible in geminally disubstituted olefins, which lack the requisite hydrogen. Feng *et al*. report a method to promote carbon migration in this class of substrates by oxidizing a catalyst intermediate to the Pd^{IV} state. The resulting pathway converts a wide variety of alkyl and aryl olefins into ketones. —JSY

Science, adg3182, this issue p. 1363

HUMAN GENOMICS

Dominance in the Biobank

According to classical genetics, traits are usually divided into dominant and recessive, with all-or-none phenotypes. Such binary traits are often observed in plant and animal breeding, but they are rarely seen in human studies outside of monogenic disorders with Mendelian inheritance patterns. Instead, many traits included in human genomic studies exhibit additive effects based on gene dosage. Palmer *et al*. examined data from the UK Biobank and identified gene loci showing dominance effects in which the phenotypes are not explained by a purely additive pattern. The authors identified 175 such loci and provided computational tools for others to perform similar analyses of dominance. They also cautioned that reliable detection of such loci requires larger sample sizes and statistical power than for purely additive effects. —YN

Science, abn8455, this issue p. 1341

NEUROSCIENCE

A nuclear exit for learning and memory

Nuclear accumulation of the second messenger cyclic adenosine monophosphate (cAMP) elicits changes in gene expression in hippocampal neurons that underlie learning and memory. Martinez *et al.* found that stimulation of the β_2-adrenergic receptor induced the nuclear export of the cAMP-degrading phosphodiesterase PDE4D5 (see the Focus by Gurevich and Gurevich). This effect required phosphorylation of the β_2-adrenergic receptor by G protein–coupled receptor kinases (GRKs). Mice expressing a mutant receptor lacking the GRK phosphorylation sites had memory deficits that were attenuated by a PDE4 inhibitor. —WW

Sci. Signal. **16**, eade3380, adg9504 (2023).

SOIL RESPIRATION

The exhalation of drylands

Semi-arid regions cause much of the interannual variability in global atmospheric carbon dioxide concentration because of seasonal fluctuations in water availability. Metz *et al*. found that recurrent carbon dioxide pulses occur at the end of the dry season over Australia, an effect that they attribute to enhanced soil respiration preceding photosynthetic uptake in semi-arid regions. The magnitudes of these pulses are two to three times larger than previous studies have inferred based on in situ measurements. —HJS

Science, add7833, this issue p. 1332

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**



PHOTO: ANDREW SYRED/SCIENCE SOURCE

PLANT SCIENCE

Splitting hairs

The leaves and shoots of tomato plants produce seven types of hairs, which can be grouped into digitate (finger-like) trichomes and peltate (shield-shaped) trichomes bearing multicellular glands. Wu *et al.* found that the concentration of a transcription factor called Woolly regulates the differentiation of trichomes. High levels of Woolly promote digitate trichome formation by up-regulating and forming a complex with other transcription factors called WOX and MYB. This module suppresses *LEAFLESS* expression, which would otherwise trigger peltate trichome development. In peltate trichomes, a negative feedback loop maintains low levels of Woolly to allow LEAFLESS activity. This work highlights how bistability arises from a concentration-dependent mechanism to influence cell fate specification. —MRS

Dev. Cell (2023) 10.1016/j.devcel.2023.01.009.

Colored scanning electron micrograph of trichomes on a tomato stem, which develop into multicellular glandular (brown) or digitate forms

PESTICIDES

Falling into trouble

The global spread of pests such as the fall armyworm is pressuring smallholder farmers in less developed countries to resort to unsafe pesticide use. Recognizing the lack of training facilities for maize-growing households in Rwanda and Uganda, Tambo *et al.* tested the effectiveness of mass media campaigns for educating farmers about the dangers of pesticides. They tested the relative value of radio broadcasts, text messaging, video screening, and plant health rallies on pest management. Questionnaires revealed that radio broadcasts were the most popular source of information, although, concerningly, the survey revealed



Crystallization in volcanic magmas, such as those of Cumbre Vieja in La Palma, Spain (pictured here), controls the rheological evolution of the melt.

VOLCANOLOGY

Figuring out lava flows

The eruption of the Cumbre Vieja volcano in La Palma, Spain, in 2021 produced fast-moving, low-viscosity lavas. Understanding the flow of these lavas requires determining crystallization during cooling because that dramatically affects their viscosity. Di Fiore *et al.* performed a series of experiments to determine the viscosity changes for different cooling rates, along with deformation experiments at constant temperatures. Their results were then compared with observations from the Tajogaite lava flow emplacement to help explain what mechanisms are responsible for the formation of features such as channels and lava tunnels. —BG

Geophys. Res. Lett. (2023) 10.1029/2022GL100970.

some use of highly dangerous pesticides and the inappropriate use of fungicides against insects. More encouragingly, radio listeners showed higher scores on pesticide knowledge questions, increased use of personal protective equipment and masks, and increased adoption (more than 27%) of nonchemical pest management practices than those who lacked access to information. —CA

J. Rural Stud. (2023) 10.1016/j.jrurstud.2023.03.001.

CANCER

Driving glioblastoma progression

Glioblastoma is a highly aggressive primary brain tumor that is initially treated by a combination of surgery, radiotherapy, and chemotherapy. However, the tumors often recur, and survival rates are poor. Hoogstrate *et al*. performed RNA sequencing on matched primary and recurrent glioblastoma resections to investigate progression, and found that tumor cell content was reduced between primary and recurrent samples. This effect correlated with increased infiltration of tumor-associated macrophages and $CD4^+$ T cells. After correcting for reduced tumor cell content, it was discovered that the expression of 722 genes was altered between primary and recurrent samples. Surprisingly, none of these were genes associated with glioblastoma tumorigenesis and progression. Clustering gene expression data revealed higher proportions of neurons and oligodendrocytes and fewer endothelial cells at recurrence, indicating that the tumor microenvironment drives glioblastoma progression. —GKA

Cancer Cell (2023) 10.1016/j.ccell.2023.02.019.

PHYSIOLOGY

Benefits after booze binge

Fibroblast growth factor 21 (FGF21) has broad-ranging effects related to alcohol use. It inhibits ethanol intake, promotes water consumption, and helps prevent liver injury. Choi *et al.* showed that FGF21 has restorative action in mice that have overimbibed. Loss of FGF21 worsened the effects of intoxication, and treatment of intoxicated mice with FGF21 helped them regain motor coordination. These effects were independent of the rate of clearance of ethanol, instead relying on the action of FGF-21 on noradrenergic neurons in the brain that influence arousal and alertness. This mechanism may have evolved to benefit animals consuming overripe fruits that have begun fermentation. —LBR

Cell Metab. (2023) 10.1016/j.cmet.2023.02.005.

MACHINE LEARNING

Tracking changes in OER oxide catalysts

Bimetallic transition metal oxides are a promising class of catalysts for the oxygen evolution reaction (OER), the key process for green hydrogen production through water electrolysis. However, their complex structures and transformations under OER limit the applicability of traditional approaches for the identification of active states and the role of different metals, which remain unclear. Timoshenko *et al.* showed that these challenges could be resolved using a combination of machine learning methods and operando quick x-ray absorption fine structure spectroscopy. Their approach successfully characterizes the structural and chemical changes in $Co_xFe_{3-x}O_4$ catalysts during OER and offers new opportunities for the rational design of efficient OER catalysts based on earth-abundant elements. —YS

J. Am. Chem. Soc. (2023) 10.1021/jacs.2c11824.

MASS MEDIA

A virulent legacy of a racist fiction

As screenings of the film *The Birth of a Nation*, a fictional portrayal of the founding of the racist hate group the Ku Klux Klan, went from theater to theater during a roadshow from 1915 to 1919, it left in its wake spikes of racist violence and long-term hatred. Drawing upon historical newspaper archives to determine where and when the film was screened, Ang shows that the 621 counties that hosted the film, across every state in the continental United States (except Kansas, which banned the film), were considerably more likely to experience a lynching or race riot in the month after the screening. Murders of minorities, but not whites, also increased. Those counties were also more likely to contain racist hate groups for decades afterward. —BW

Am. Econ. Rev. (2023) 10.1257/aer.20201867.

PHOTO: IMV/GETTY IMAGES

Check for updates

RESEARCH ARTICLE SUMMARY

NEUROIMMUNOLOGY

Multimodal control of dendritic cell functions by nociceptors

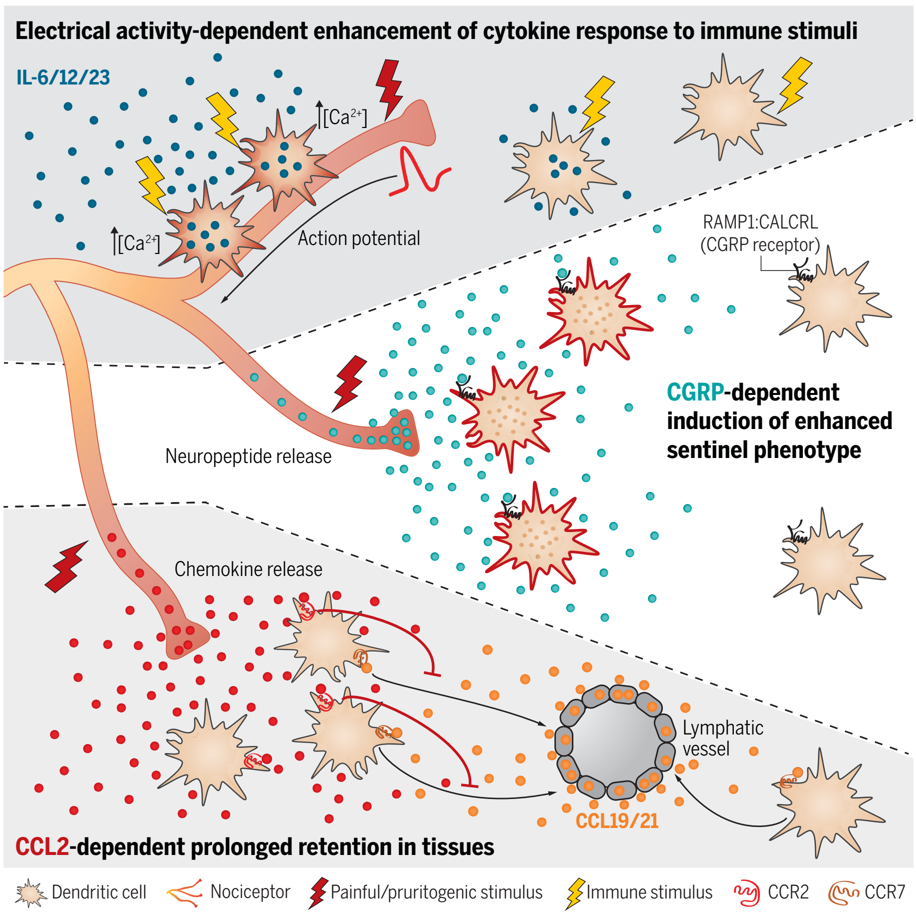
Pavel Hanč, Rodrigo J. Gonzalez, Irina B. Mazo, Yidi Wang, Talley Lambert, Gloria Ortiz, Evan W. Miller, Ulrich H. von Andrian*

INTRODUCTION: There is growing evidence that neuroimmune interactions regulate immune and inflammatory responses. In particular, nociceptors, the afferent neurons that transmit the sensation of pain or itch in response to noxious stimuli, modulate immune cell functions in the peripheral tissues that they innervate. One of the major nociceptor targets are dendritic cells (DCs), which orchestrate local inflammatory responses and control adaptive immunity. Although nociceptors have been shown to exert control over DCs in several pathophysiological settings, the molecular mechanisms underlying nociceptor–DC cross-talk and the full scope of nociceptor involvement in DC immunobiology remain to be established.

RATIONALE: To determine the communication framework between nociceptors and DCs, we established an in vitro coculture system for the two cell types that enabled us to interrogate bona fide nociceptor–DC interactions without the involvement of other cell types. We explored the impact of nociceptors on both activated and steady-state DCs and defined, at the molecular level, communication pathways that rely on either physical cell–cell contact or nociceptor-derived soluble mediators. Subsequently, we investigated whether key observations in our reductionist in vitro system were predictive of the nociceptor–DC dialogue in several in vivo scenarios.

RESULTS: We identified three molecularly distinct modalities of communication used by nociceptors to control DC functions in a context-dependent manner. First, through the production of the chemokine CCL2, nociceptors can attract DCs and regulate their tissue dwell time. Second, through the activation-induced release of the neuropeptide calcitonin gene–related peptide (CGRP), nociceptors induce a transcriptional program in DCs characterized by the expression of *Il1b* as well as multiple other genes important for pathogen resistance and sentinel functions, but no overt DC activation. Third, through direct electrical coupling, firing nociceptors trigger a Ca^{2+} flux and membrane depolarization in DCs and potentiate DC responses to inflammatory stimuli such as Toll-like receptor (TLR) agonists. Accordingly, in vivo activation of dermal nociceptors enhanced the sentinel phenotype of DCs in the absence of immune stimuli and amplified DC-dependent tissue inflammation in response to TLR agonists. Furthermore, conditional ablation of CCL2 in nociceptors compromised the ability of dermal DCs to amplify local inflammation in the skin and to initiate adaptive immune responses against skin-derived antigens.

CONCLUSION: The multiple communication modalities characterized here reveal how nociceptors and DCs form a neuroimmune unit that integrates the nociceptors' rapid responsiveness and finely honed ability to sense and respond to noxious stimuli with the DCs' ability to coordinate innate and adaptive immune responses. This merger of unique capabilities allows the nociceptor–DC unit to act as an advanced warning system. DCs in barrier tissues are therefore a priori primed to anticipate imminent pathogen encounter, respond more vigorously upon painful pathogen exposure, and fine-tune the subsequent orchestration of adaptive immune responses to peripheral antigens. ■



Nociceptors communicate with dendritic cells (DCs) using three distinct pathways to elicit context-dependent responses. (i) The release of CCL2 attracts DCs and regulates their dwell time within barrier tissues and antigen transport to lymph nodes; (ii) pain-induced release of the neuropeptide CGRP alters the transcriptional profile of steady-state DCs to assume enhanced sentinel functions; and (iii) when activated DCs contact firing nociceptors, the electrical activity amplifies their proinflammatory cytokine response. IL-6/12/23, interleukin-6, interleukin-12, interleukin-23.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: uva@hms.harvard.edu
Cite this article as P. Hanč *et al.*, *Science* **379**, eabm5658 (2023). DOI: 10.1126/science.abm5658

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm5658

RESEARCH ARTICLE

NEUROIMMUNOLOGY

Multimodal control of dendritic cell functions by nociceptors

Pavel Hanč[1,2], Rodrigo J. Gonzalez[1,2], Irina B. Mazo[1,2], Yidi Wang[1,2], Talley Lambert[3], Gloria Ortiz[4], Evan W. Miller[4], Ulrich H. von Andrian[1,2]*

It is known that interactions between nociceptors and dendritic cells (DCs) can modulate immune responses in barrier tissues. However, our understanding of the underlying communication frameworks remains rudimentary. Here, we show that nociceptors control DCs in three molecularly distinct ways. First, nociceptors release the calcitonin gene–related peptide that imparts a distinct transcriptional profile on steady-state DCs characterized by expression of pro–interleukin-1β and other genes implicated in DC sentinel functions. Second, nociceptor activation induces contact-dependent calcium fluxes and membrane depolarization in DCs and enhances their production of proinflammatory cytokines when stimulated. Finally, nociceptor-derived chemokine CCL2 contributes to the orchestration of DC-dependent local inflammation and the induction of adaptive responses against skin-acquired antigens. Thus, the combined actions of nociceptor-derived chemokines, neuropeptides, and electrical activity fine-tune DC responses in barrier tissues.

Nociceptors are sensory neurons that respond to noxious stimuli by eliciting sensations of pain or itch and by modulating immune responses (*1*–*4*). Their somata reside in dorsal root ganglia (DRGs) with axons projecting to both the spinal cord and peripheral tissues (*5*). Nociceptors densely innervate barrier tissues (*6*) such as the skin and mucosal surfaces, which are also populated by dendritic cells (DCs). These sentinel leukocytes sense pathogen- and danger-associated molecular patterns and play critical roles in antigen presentation and control of adaptive immunity. Moreover, DCs coordinate local inflammation by secreting cytokines and other mediators (*7*). Intravital imaging demonstrated that dermal DCs congregate tightly around nociceptors, which profoundly affect DC functions (*8*–*10*). However, although these observations suggest direct cell–cell communication, the requirement for physical DC–nociceptor interactions versus a possible involvement of other intermediary cells has not been established. Furthermore, the role of nociceptor-derived neuropeptides in DC–nociceptor communication has been controversial, because neuropeptides appear dispensable in some experimental settings but exert either pro- or anti-inflammatory effects in other models (*8*, *9*, *11*).

[1]Department of Immunology, Harvard Medical School, Boston, MA 02115, USA. [2]The Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA 02139, USA. [3]Cell Biology Microscopy Facility, Harvard Medical School, Boston, MA 02115, USA. [4]Departments of Chemistry, Molecular & Cell Biology, and Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720, USA.
*Corresponding author. Email: uva@hms.harvard.edu

Nociceptors potentiate cytokine responses by DCs in vitro

To address these outstanding questions, we cocultured primary DRG neurons and Flt3L-induced bone marrow DCs (BMDCs) (fig. S1A). Neuronal cultures contained >90% small-diameter NaV1.8$^+$ nociceptors (fig. S1B), whereas BMDC cultures comprised three major subsets: type 1 and 2 conventional DCs (cDCs) and plasmacytoid DCs (pDCs) (fig. S1C). Using this system, we examined DC responses to a Toll-like receptor 7 (TLR7) agonist, imiquimod (IMQ). Nociceptors did not affect DC maturation marker expression upon IMQ treatment (fig. S2A), but their presence significantly enhanced DC production of interleukin-12 (IL-12) p40, the shared subunit of IL-12 and IL-23 (Fig. 1A), as well as IL-6 (Fig. 1B). Conversely, production of another proinflammatory cytokine, tumor necrosis factor α (TNFα), was reduced (Fig. 1C). A likely mechanism for TNFα down-regulation is the action of calcitonin gene–related peptide (CGRP), a nociceptor-derived neuropeptide known to inhibit cytokine production by myeloid leukocytes (*11*). Indeed, olcegepant, a CGRP receptor antagonist, rescued TNFα production in DC–nociceptor cocultures, but had no impact on DCs alone (fig. S2B). Olcegepant did not affect DC production of IL-12 p40 or IL-6 regardless of whether nociceptors were present (fig. S2, C and D), indicating that this cytokine-promoting effect of nociceptors on DCs was CGRP independent. By contrast, a different myeloid cell type, BM-derived macrophages, showed decreased or unaltered cytokine responses to IMQ in nociceptor cocultures (Fig. 1D). Finally, coculture with hippocampal or cortical neurons did not alter DC responses to IMQ (Fig. 1E and fig. S2E). Thus, the potentiation of inflammatory cytokine production (other than TNFα) is a unique and specific consequence of the DC–nociceptor dialogue.

Numerous stimuli can trigger inflammatory cytokine production by DCs, including damage-associated molecular patterns (DAMPs) released from dead or distressed cells (*12*). A DC response to DAMPs from dead nociceptors could conceivably provide a trivial explanation for the exacerbated IMQ effect. However, DCs cocultured with nociceptors killed by prior fixation showed no enhanced cytokine production (Fig. 1F and fig. S2F). Moreover, treatment with lidocaine (a voltage-gated sodium channel blocker) and QX314 (a membrane-impermeable quaternary derivative of lidocaine), which together produce a long-lasting blockade of neuronal activity (*13*), dampened the ability of nociceptors to potentiate DC cytokine responses (Fig. 1G and fig. S2G). Thus, nociceptors must be alive and electrically active to exert their effects.

IMQ stimulated IL-12 p40 production in only a fraction of DCs, which increased in a concentration-dependent manner. Nociceptor cocultures increased the frequency of responding cDCs, but did not grossly alter the amount of cytokine detected per cell (Fig. 1H). By contrast, IMQ treatment of pDCs induced only a modest response, which was minimally affected by nociceptors (fig. S3A). Accordingly, the accumulation of interferon β (IFN-β), a prototypical pDC cytokine (*14*), was also unaffected by nociceptors (fig. S3B). Finally, although only cDC2s and pDCs produced IL-6 in response to IMQ (fig. S3C), the proportion of IL-6$^+$ cDC2s was significantly increased in the presence of nociceptors (fig. S3C), whereas pDCs in nociceptor cocultures showed only a weak, statistically nonsignificant trend toward increased IL-6 production (fig. S3C). Thus, cDCs, but not pDCs, are affected by nociceptors, at least under these experimental conditions, and the presence of electrically active nociceptors increases the proportion of DCs responding to a given dose of IMQ, presumably by decreasing the activation threshold to initiate cytokine production.

Cytokine potentiation requires concomitant activation of nociceptors and DCs

The principal IMQ receptor, TLR7 (*15*), is robustly expressed in murine DCs, but it has also been reported in sensory neurons (*16*). To clarify its role in this process, we established cocultures in which either nociceptors or DCs had been harvested from *Tlr7*$^{-/-}$ mice (*17*). Cytokine responses to IMQ were abrogated when DCs lacked TLR7, whereas *Tlr7*$^{-/-}$ and wild-type (WT) nociceptors equally enhanced cytokine production by WT DCs (Fig. 1I and

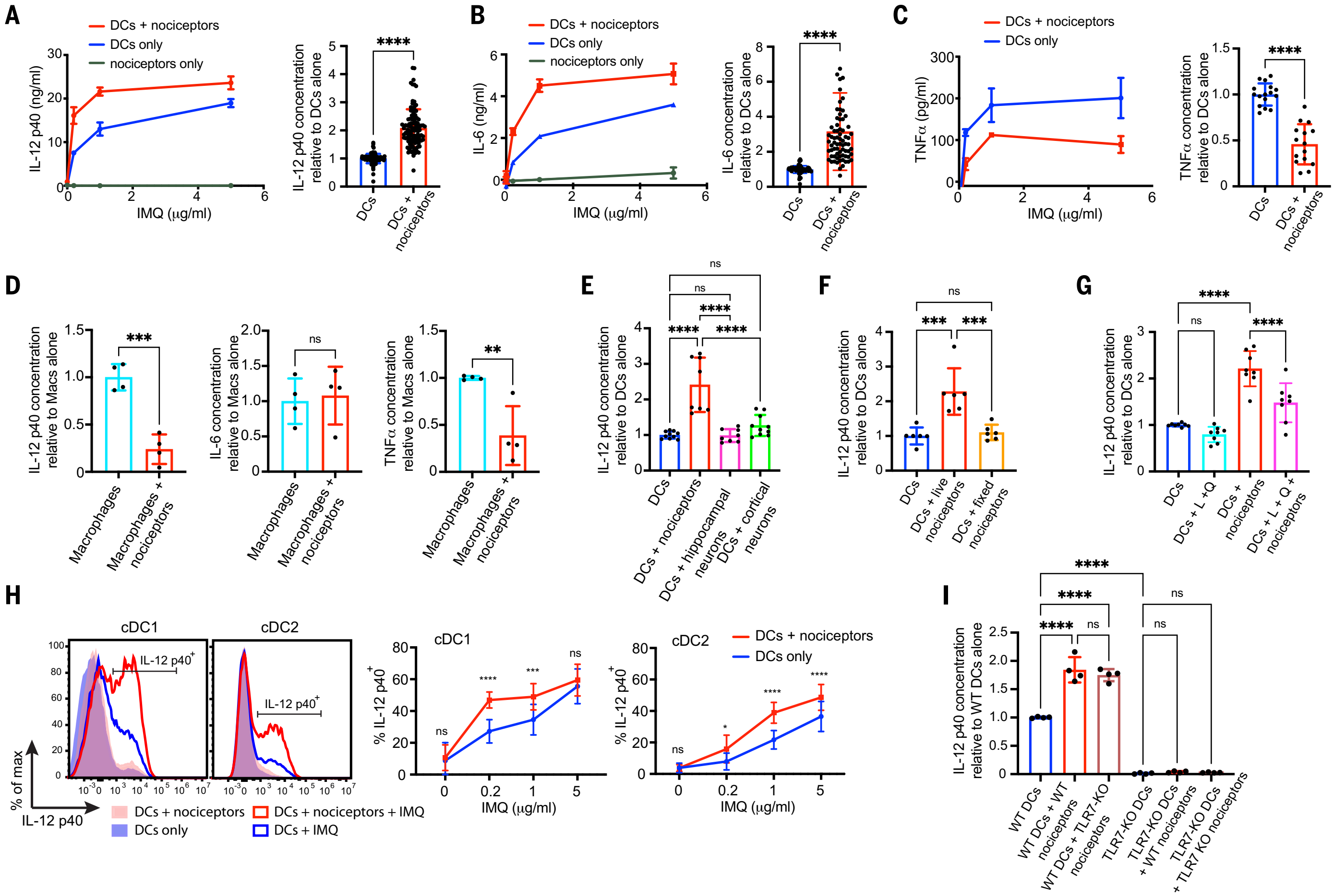


Fig. 1. Nociceptors enhance proinflammatory cytokine production by IMQ-activated DCs in vitro. BMDCs and nociceptors were cultured either separately or together and treated overnight with proinflammatory stimuli. Indicated cytokines were measured by ELISA in supernatants. (**A** to **C**) One representative dose response to IMQ (left panels) and summary of 42 (n = 104) (A), 37 (n = 74) (B), and 7 (n = 16) (C) experiments (right panels) for 1 μg/ml IMQ. (**D**) BM-macrophages were cultured alone or in the presence of nociceptors and treated with 1 μg/ml IMQ overnight. Summary of two experiments (n = 4) is shown. (**E**) BMDCs were cocultured with nociceptors, cortical neurons, or hippocampal neurons and treated with 1 μg/ml IMQ overnight. Summary of three experiments (n = 6 to 10) is shown. (**F**) BMDCs were cocultured with live or fixed nociceptors and treated with 1 μg/ml IMQ. Summary of three experiments (n = 6) is shown. (**G**) DCs were cocultured with nociceptors and treated with lidocaine + QX314 and 1 μg/ml IMQ. Summary of four experiments (n = 8) is shown. (**H**) Intracellular content of IL-12 p40 in BMDCs left untreated (full histograms) or treated with IMQ (open histograms) in isolation or in a coculture with nociceptors was assessed using flow cytometry. One representative experiment (left) and quantitation of eight independent experiments (right; n = 16) are shown. (**I**) WT or *Tlr7*-knockout (KO) DCs were cocultured with WT or *Tlr7*-KO nociceptors and treated with 1 μg/ml IMQ. Summary of two (n = 4) experiments is shown. Across all panels, data represent mean ± SD. Unpaired t test [(A) to (D)], one-way ANOVA with Tukey's multiple-comparisons test (F), or two-way ANOVA with Tukey's multiple-comparisons test [(E) and (G) to (I)] were used for statistical analysis: *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

fig. S4A). Thus, nociceptors cannot by themselves elicit a cytokine response in DCs unless DCs are directly stimulated. Nonetheless, concomitant IMQ stimulation of nociceptors was required, but independently of TLR7. IMQ can directly gate the cation channel TRPA1 (*18*, *19*), and a TRPA1 inhibitor prevented nociceptor enhancement of DC responses to IMQ (fig. S4B).

The fact that IMQ may act on both DCs and nociceptors although through different mechanisms prompted us to investigate other immune stimuli, including zymosan (a TLR2 and Dectin-1 agonist), polyinosinic:polycytidylic acid (poly-I:C) (a TLR3 agonist), lipopolysaccharide (LPS) (a TLR4 agonist), flagellin (a TLR5 agonist), or CpG (a TLR9 agonist). All of these agents activate DCs, but only LPS, flagellin, and zymosan stimulate nociceptors by direct gating of ion channels or through membrane receptors such as TLR4 or Dectin-1 (*20*–*22*). Accordingly, nociceptors potentiated DC cytokine responses to zymosan, LPS, and flagellin, but not to poly-I:C or CpG (Fig. 2A and fig. S4C). Finally, as expected, nociceptors potentiated the LPS response irrespective of TLR7 expression by DCs (fig. S4D), indicating that $Tlr7^{-/-}$ DCs did not respond to IMQ because of their inability to register the stimulus and not because of other inherent functional defects.

Next, to test whether nociceptors can also enhance DC cytokine responses to infectious microbes, we exposed nociceptor–DC cocultures to viral [influenza A virus (IAV)], bacterial (ultraviolet light–inactivated *Streptococcus pnneumoniae*), or fungal (*Candida albicans*) pathogens that are commonly encountered in barrier tissues. IL-12 p40 was induced by every tested pathogen, whereas IL-6 was only induced by *S. pneumoniae* and IAV. Regardless of the microbial challenge, coculture with nociceptors markedly enhanced these responses (Fig. 2B and fig. S4E), suggesting that the paradigm of nociceptor enhancement of DC cytokine production may

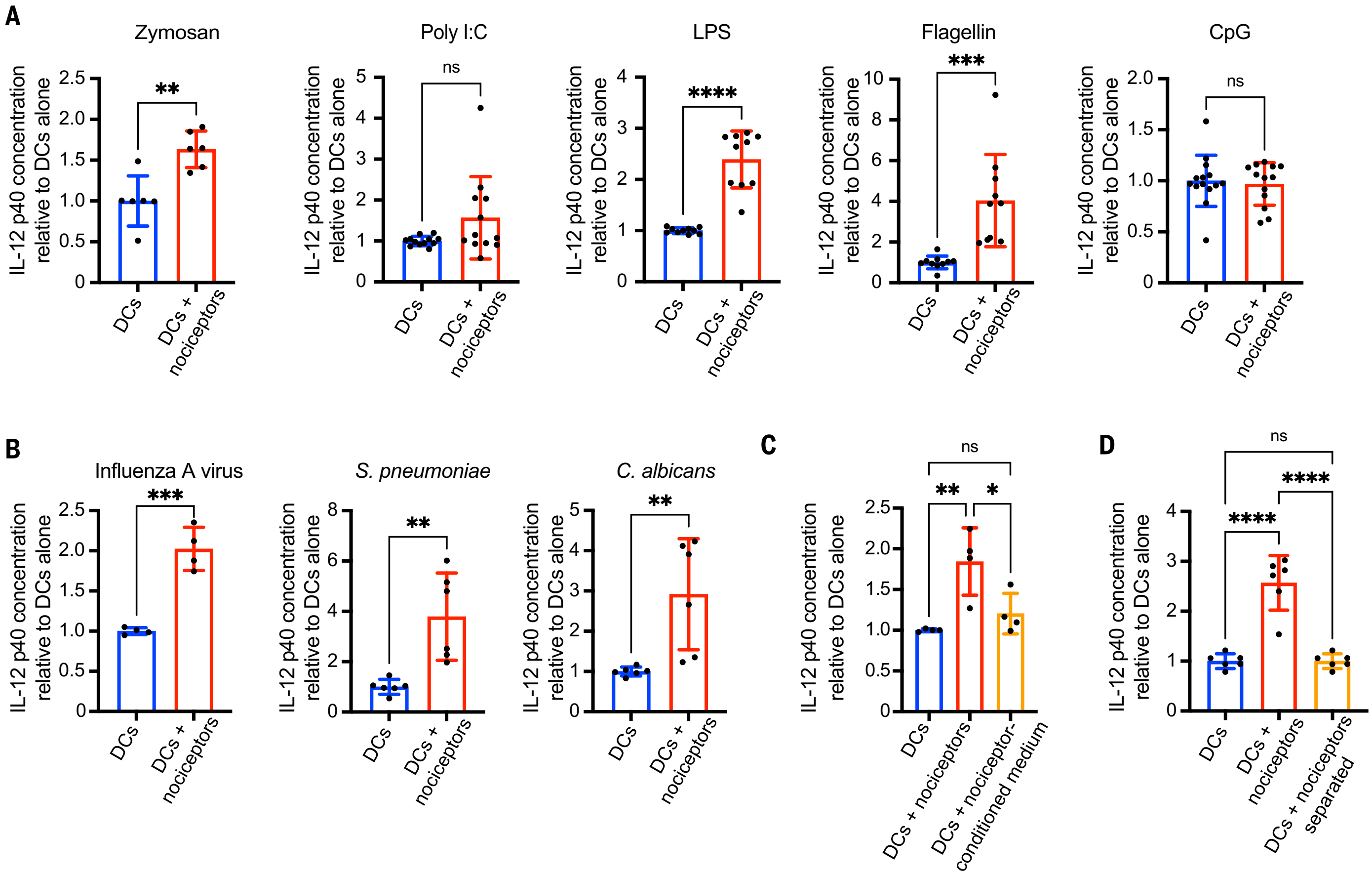


Fig. 2. Nociceptors enhance DC cytokine response to various stimuli in a contact-dependent manner. (**A** and **B**) BMDCs and nociceptors were either cultured separately or together and treated overnight with indicated nonmicrobial (A) or microbial (B) stimuli. Summary of two to five experiments (n = 4 to 12) is shown. (**C**) BMDCs were cultured alone, with nociceptors, or in a transferred medium in which nociceptors had been stimulated and treated with IMQ overnight. Summary of three experiments (n = 4) is shown. (**D**) BMDCs were cultured in Transwell plates either alone or with nociceptors in a manner that allowed or prevented direct physical contact between the cell types. Summary of three experiments (n = 6) is shown. IL-12 p40 concentration was measured by ELISA in supernatants. Across all panels, data represent mean ± SD. Unpaired t test [(A) and (B)] or one-way ANOVA with Tukey's multiple-comparisons test [(C) and (D)] were used for statistical analysis: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

apply to a host of inflammatory and infectious stimuli.

DC cytokine potentiation by nociceptors requires physical contact but not neuropeptides

Upon activation, nociceptors release immunomodulatory neuropeptides, particularly the CGRP-family of neuropeptides (CGRP, adrenomedullin, and intermedin) (23), substance P (SP) (24), vasoactive intestinal peptide (VIP) (25), and pituitary adenylate cyclase-activating peptide (PACAP) (25). However, DC exposure to neuropeptides did not enhance the IMQ-induced production of IL-12 p40 or IL-6, nor did neuropeptides induce de novo cytokine responses in unstimulated DCs (fig. S4F). Similarly, nociceptor-conditioned medium did not enhance cytokine production by IMQ-stimulated DCs, suggesting that direct cell–cell contact may be required (Fig. 2C and fig. S4G). Indeed, when DCs and nociceptors were separated by a filter membrane in a Transwell device that allowed diffusion of soluble molecules while preventing physical contact, DCs produced no more IL-12 p40 than did DCs cultured alone (Fig. 2D). Furthermore, the accumulation of IL-6 was decreased, albeit not quite to the levels observed in DC-only cultures (fig. S4H). Thus, the presence of electrically competent nociceptors enhances DC cytokine responses to multiple exogenous immune stimuli in a manner independent of neuropeptides and is dependent on physical contact or at least close physical proximity.

Nociceptors attract DCs by secreting CCL2

Seeing the importance of physical contact for nociceptor–DC communication in vitro, and in light of their close proximity in vivo (9, 10), we investigated whether nociceptors produce DC chemoattractants. Indeed, all DC subsets migrated toward nociceptor-conditioned medium in a concentration-dependent manner (Fig. 3A and fig. S5, A and B). Using two multiplexed assays, we identified several known chemoattractants in nociceptor-conditioned medium, with CCL2 and chemerin being the most abundant (fig. S5, C and D). Although chemerin immunodepletion had no effect, CCL2 depletion significantly decreased DC migration toward nociceptor-conditioned medium (fig. S5E). Furthermore, medium conditioned by nociceptors from *Ccl2*$^{-/-}$ mice (26) showed a reduced capacity to attract DCs (Fig. 3A and fig. S5, A and B), indicating that CCL2 is the predominant DC chemoattractant secreted by nociceptors. Accordingly, CCL2 was detectable in medium conditioned by WT but not *Ccl2*$^{-/-}$ nociceptors (fig. S5F) and increased after nociceptor stimulation with capsaicin or IMQ (fig. S5G). A previous report linked CCL2 up-regulation in nociceptors to activation of

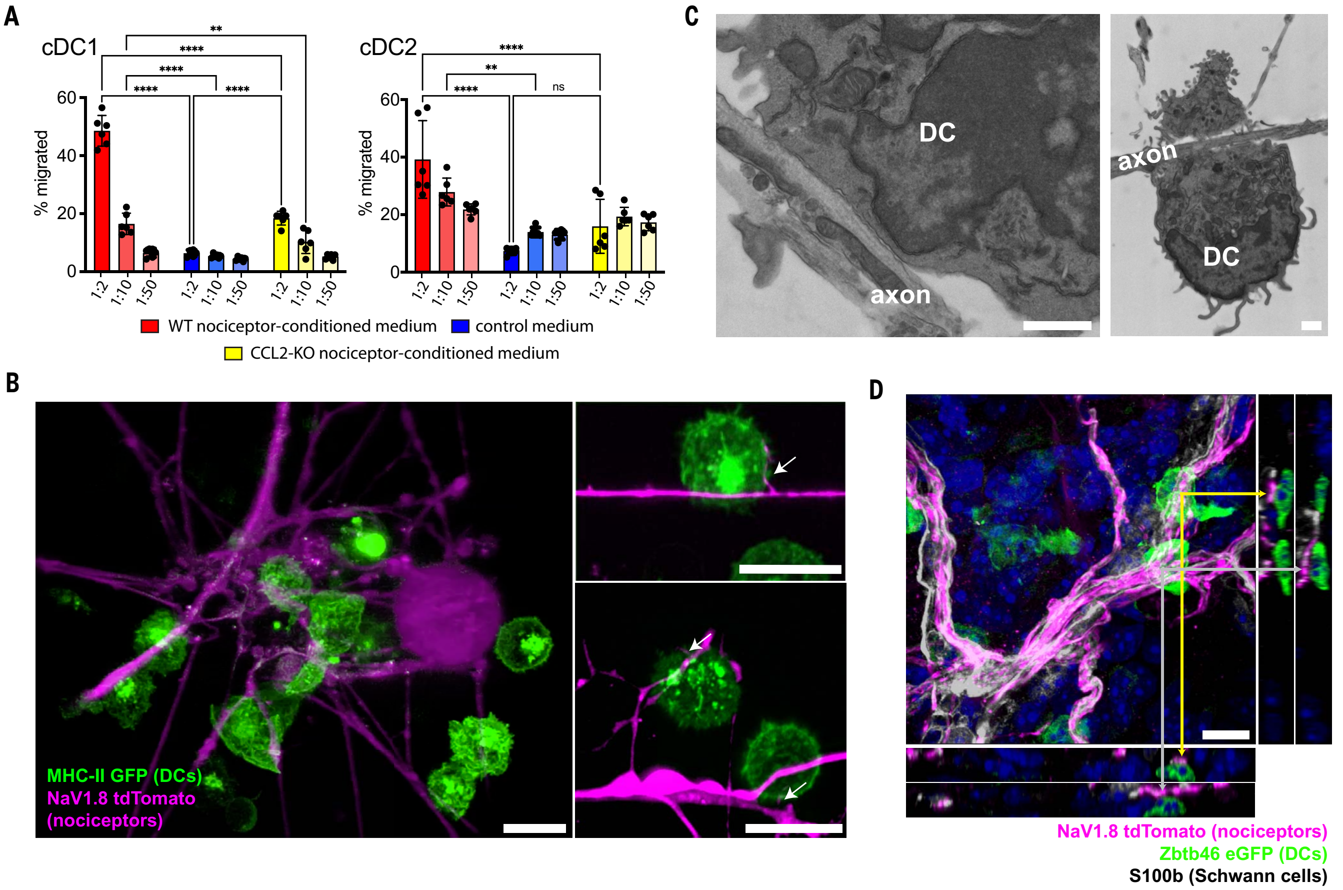


Fig. 3. Nociceptors attract and physically interact with DCs in vitro. (**A**) Migration of BMDCs toward WT nociceptor-conditioned, control, and *Ccl2*$^{-/-}$ nociceptor-conditioned medium was assessed in a Transwell chemotaxis assay (5-μm pore size). Summary of three experiments (n = 6) is shown. Data represent mean ± SD. Two-way ANOVA with Tukey's multiple-comparisons test was used for statistical analysis: *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. (**B**) DCs from MHC-II eGFP mice were cocultured with nociceptors from NaV1.8$^{Cre/+}$ ROSA$^{tdTomato/+}$ donors and imaged using a lattice light-sheet microscope (see also movies S2 and S3). Arrowheads highlight some of the points of contact between nociceptors and DCs. Scale bars, 10 μM. (**C**) Transmission electron micrographs of a DC contacting a nociceptor axon. Two representatives of >10 similar images are shown. Scale bars, 1 μM. (**D**) Dermal sheets from BM chimeric animals expressing tdTomato under the NaV1.8 promoter in somatic cells and eGFP under the Zbtb46 promoter in the hematopoietic lineage were stained with anti-S100 antibody to identify Schwann cells and analyzed using confocal microscopy.

MyD88, a signaling adaptor critical for TLR function (*27*). Indeed, LPS-induced ligation of TLR4 resulted in markedly increased levels of CCL2 (fig. S5H). Moreover, in agreement with previous studies, unstimulated cDC1s and cDC2s expressed uniformly high levels of the CCL2 receptor CCR2, which was down-regulated upon LPS exposure, coincident with CCR7 up-regulation (fig. S5I) (*28*). In macrophages, CCL2 reportedly affects the secretion of cytokines including IL-12 (*29*). However, CCL2 had no effect on cytokine production by DCs, because *Ccl2*$^{-/-}$ and WT nociceptors potentiated the DC response to IMQ equally (fig. S5J).

DCs engage in tight and dynamic interactions with nociceptors

Next, using live-cell imaging, we observed that DCs engaged in tight and highly dynamic physical interactions with nociceptors (Fig. 3B, fig. S6A, and movies S1 to S3). Furthermore, transmission electron microscopy revealed a tight apposition of the plasma membranes of interacting cells (Fig. 3C). To determine whether similarly tight contacts occur in vivo, we transplanted BM from *Zbtb46*^{eGFP} (where eGFP is enhanced green fluorescent protein) donors (*30*) into irradiated *Scn10a*$^{Cre/+}$*Rosa26*$^{tdTomato/+}$ recipients. Nociceptors and DCs in these chimeras were identifiable by expression of tdTomato and GFP, respectively, and confocal microscopy analysis of dermal sheets revealed physical contacts between them (Fig. 3D).

Nociceptor activation induces calcium flux in interacting DCs

Given the importance of physical contact for effective communication between nociceptors and DCs and the observation that nociceptors attract DCs by the secretion of CCL2, we performed calcium imaging to assess whether and how signals propagate from activated neurons to interacting DCs. Upon treatment with the TRPV1 channel agonist capsaicin (*31*), a rapid increase in intracellular calcium occurred in both nociceptors and DCs (Fig. 4, A and B, and movie S4). Capsaicin did not elicit a calcium flux in DCs alone (Fig. 4B and movie S5), indicating that the capsaicin effect in cocultures was caused by the activation of nociceptors, not DCs.

To better approximate the anatomic relationship between DCs and nociceptors, we next used a culture device with two compartments connected by fluidically isolated microgrooves (*32*) (fig. S6B). Nociceptors plated in one compartment, the "neuronal compartment," sprouted axons across the microgrooves into the second compartment, the "DC compartment." Capsaicin stimulation of nociceptors in the neuronal compartment triggered axonal action potential propagation to the DC

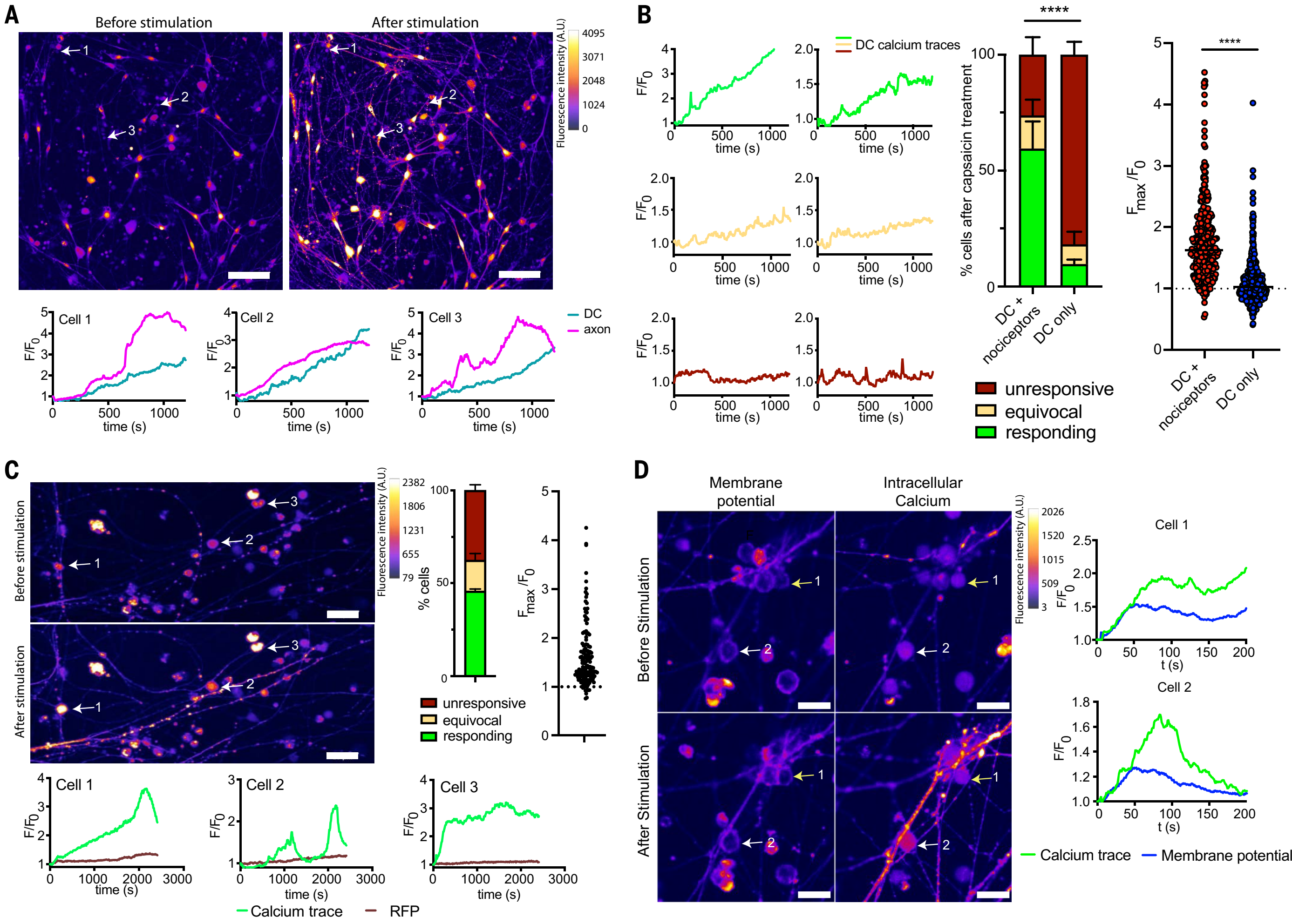


Fig. 4. Activation of nociceptors induces calcium mobilization and membrane depolarization in DCs. (**A**) Cells in nociceptor:DC cocultures were loaded with the calcium indicator dye Fluo-4 and imaged on a spinning disk confocal microscope. Left panel shows the culture before and right panel after the addition of capsaicin. Representative calcium traces of DCs and proximal axons (arrowheads) are shown in bottom panels (see also movies S4 and S5). (**B**) Calcium traces of representative responder (green), equivocal (yellow), and unresponsive (dark red) DCs (left panel) and quantification of four independent experiments as shown in (A), comparing the proportion of responder cells (middle panel) and the magnitude of calcium response in each cell (right panel). (**C**) Nociceptors were plated in one compartment (neuronal compartment) of a microfluidic device consisting of two wells that were separated by fluidically-isolated microgrooves. Once neuronal axons had grown across the microgrooves, DCs were added to the second compartment (DC compartment). All cells were loaded with Fluo-4, and the DC compartment was imaged on a spinning disk confocal microscope after capsaicin addition to the neuronal compartment. Some of the responding DCs are highlighted, and their calcium traces are shown in the lower panels. One representative and a quantitation of three independent experiments are shown (see also movie S6). (**D**) Cells in a nociceptor–DC coculture were loaded with Fluo-4 and a membrane potential indicator dye (BeRST) and imaged on a spinning disk confocal microscope. Top panels show the culture before and bottom panels after the addition of capsaicin. Some of the responding DCs are highlighted, and their calcium and membrane potential traces are shown in the right panel. One representative of two experiments is shown (see also movie S10). Across all imaging panels, warmer colors represent higher intracellular calcium concentration except in (D), left panels, where they indicate lower membrane potential (depolarization).

compartment, where DCs interacting with firing axonal segments fluxed calcium (Fig. 4C and movie S6) comparably to conventional cocultures (fig. S6C). Although some DCs fluxed calcium in sync with contacting axons (fig. S6D and movie S7), calcium flux kinetics in the DC population as a whole were nonuniform (Fig. 4C and fig. S6E), a phenomenon likely attributable to the dynamic nature of interactions, asynchronous firing of different axons, nonhomogeneous distribution of TRPV-1 expression among nociceptors, and potentially also differences between DC subsets.

Nociceptor activation induces membrane depolarization in interacting DCs

Neuronal calcium influx is triggered by membrane depolarization as action potentials propagate along axons. Therefore, we wondered whether the calcium flux in DCs was also accompanied by membrane depolarization. Indeed, using a ratiometric fluorescent probe, Di-8-ANNEPS, capsaicin-induced membrane depolarization was detected in both cell types in DC–nociceptor cocultures. Capsaicin, however, had no effect on DCs alone (fig. S6F and movies S8 and S9). Similarly, DCs and nociceptors stained with BeRST, a voltage-sensitive dye compatible with calcium imaging (*33*), depolarized and fluxed calcium in response to capsaicin (Fig. 4D and movie S10).

Thus, nociceptors attract DCs by producing CCL2 and engage in tight physical interactions,

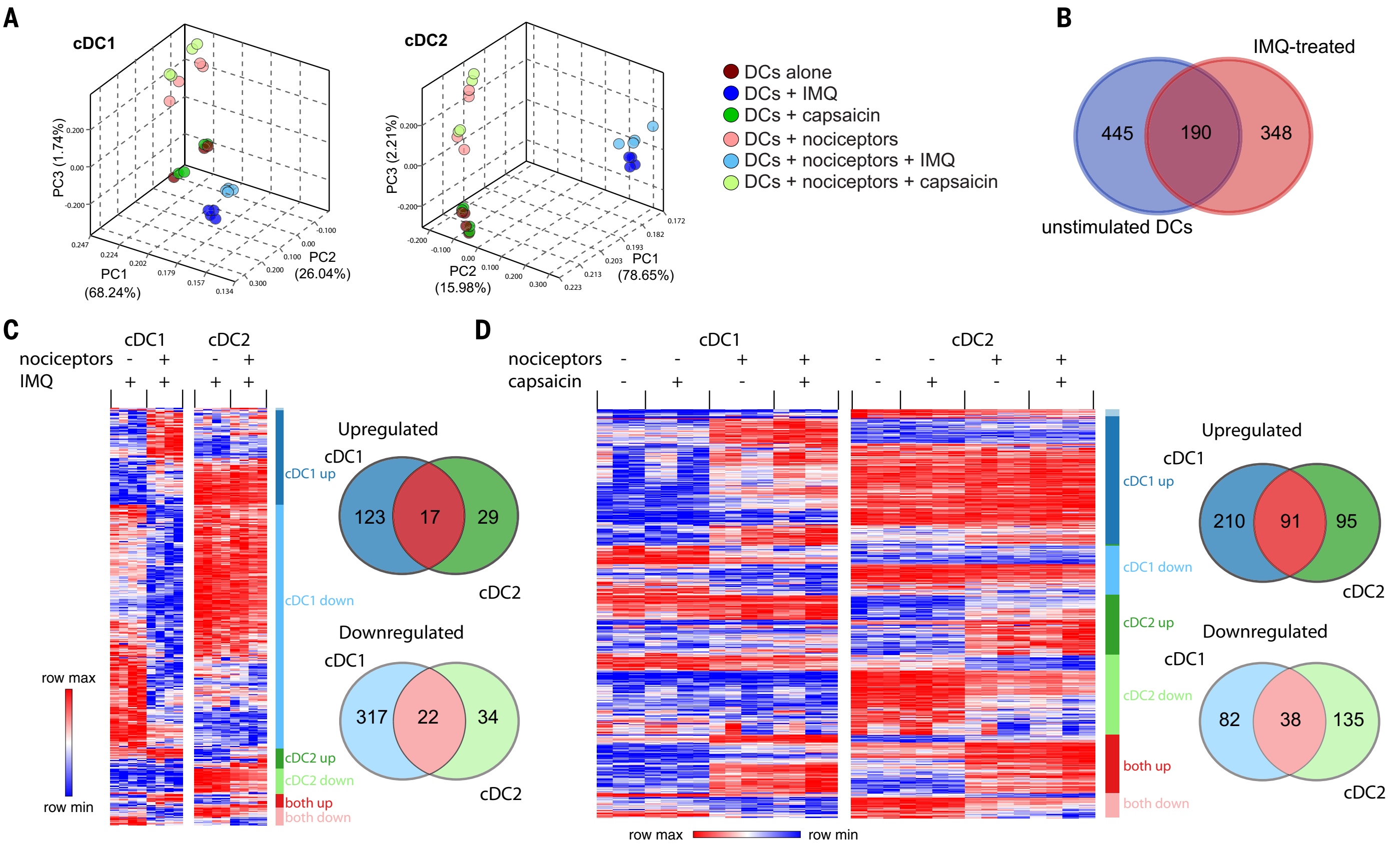


Fig. 5. Nociceptors modulate the DC transcriptome. (**A**) PCA of RNA-seq data from DCs treated with IMQ or capsaicin in the presence or absence of nociceptors. (**B**) Venn diagram of differentially regulated genes across DC subsets in coculture with nociceptors compared with DC monoculture. (**C** and **D**) Heatmap of genes differentially regulated (FC ≥ 2, FDR = 0.1) between DCs stimulated with IMQ in monoculture and in coculture with nociceptors (C) or DCs left untreated or treated with capsaicin in the presence or absence of nociceptors (D). Colored bars and sections of the Venn diagrams designate gene sets that are uniquely differentially regulated in cDC1s (blue), cDC2s (green), or co-regulated in the two subsets (red). Darker shades correspond to up-regulated genes and lighter shades to down-regulated genes.

presumably mediated by a yet-to-be-identified adhesion molecule(s). Electrical nociceptor activity induces in interacting DCs membrane depolarization and calcium influx, which may modulate myeloid cell functions (*34*, *35*). Indeed, calcium mobilization has been shown to increase IL-12 production by DCs upon TLR stimulation (*36*). Accordingly, coadministration of a calcium ionophore boosted the DC response to IMQ (fig. S7, A and B), analogous to our observations in nociceptor–DC cocultures.

Nociceptors induce transcriptomic changes in DCs

Next, to assess the impact of nociceptors beyond amplifying DC cytokine responses, we performed RNA sequencing (RNA-seq) on DCs cultured in the presence or absence of nociceptors and with or without IMQ or capsaicin. Principal components analysis (PCA) revealed that the mere presence of nociceptors induced profound transcriptomic changes in both resting and IMQ-treated cDCs (Fig. 5A). Further marginal changes observable in PCA space were induced by capsaicin stimulation of cocultures, whereas isolated DCs showed altered gene expression profiles only after exposure to IMQ, but not capsaicin. Across all conditions, 983 genes in cDCs were differentially regulated by nociceptors with a fold change (FC) ≥ 2 [false discovery rate (FDR) = 0.1]. Of the differentially regulated genes, a minority (190 genes) were shared between the IMQ-treated and -untreated groups, indicative of two largely independently regulated genesets, one modulated by nociceptors in immature DCs (445 genes) and a second that only became apparent upon DC activation (348 genes) (Fig. 5, B to D, and fig. S8A). Compared with cocultures with unstimulated nociceptors, capsaicin exposure resulted in only 16 differentially regulated genes in cDCs (FC ≥ 2; FDR = 0.1) (fig. S8B), suggesting that nociceptors express the most relevant communication molecule(s) constitutively, at least in vitro. Finally, consistent with our earlier observations (fig. S3, A to C), the presence of nociceptors exerted minimal effects on pDCs, in which only 38 genes were differentially expressed across all conditions (fig. S8, C and D).

Of the differentially regulated genes in IMQ-activated cDCs, most were selectively regulated in cDC1s, with a minority shared or unique to cDC2s (Fig. 5C). Gene-set enrichment analysis (GSEA) (*37*, *38*) revealed that among the hallmark gene sets (*39*), the most prominently upregulated pathways in activated cDCs exposed to nociceptors were the glycolysis and hypoxia response pathways (fig. S9, A to C). This was consistent with the increased proportion of cytokine-producing cells (Fig. 1H and fig. S3C), because these two gene sets correlate with DC activation status (*40*). Additionally, GSEA revealed a down-regulation in the unfolded protein response and apoptotic signaling in activated cDC1s, suggesting that nociceptors may improve activated cDC1 survival. Most of the differentially regulated genes in immature cDCs (Fig. 5D) were likewise regulated in a subset-specific manner. However, the number of genes was comparable between cDC subsets, and more genes were co-regulated. These

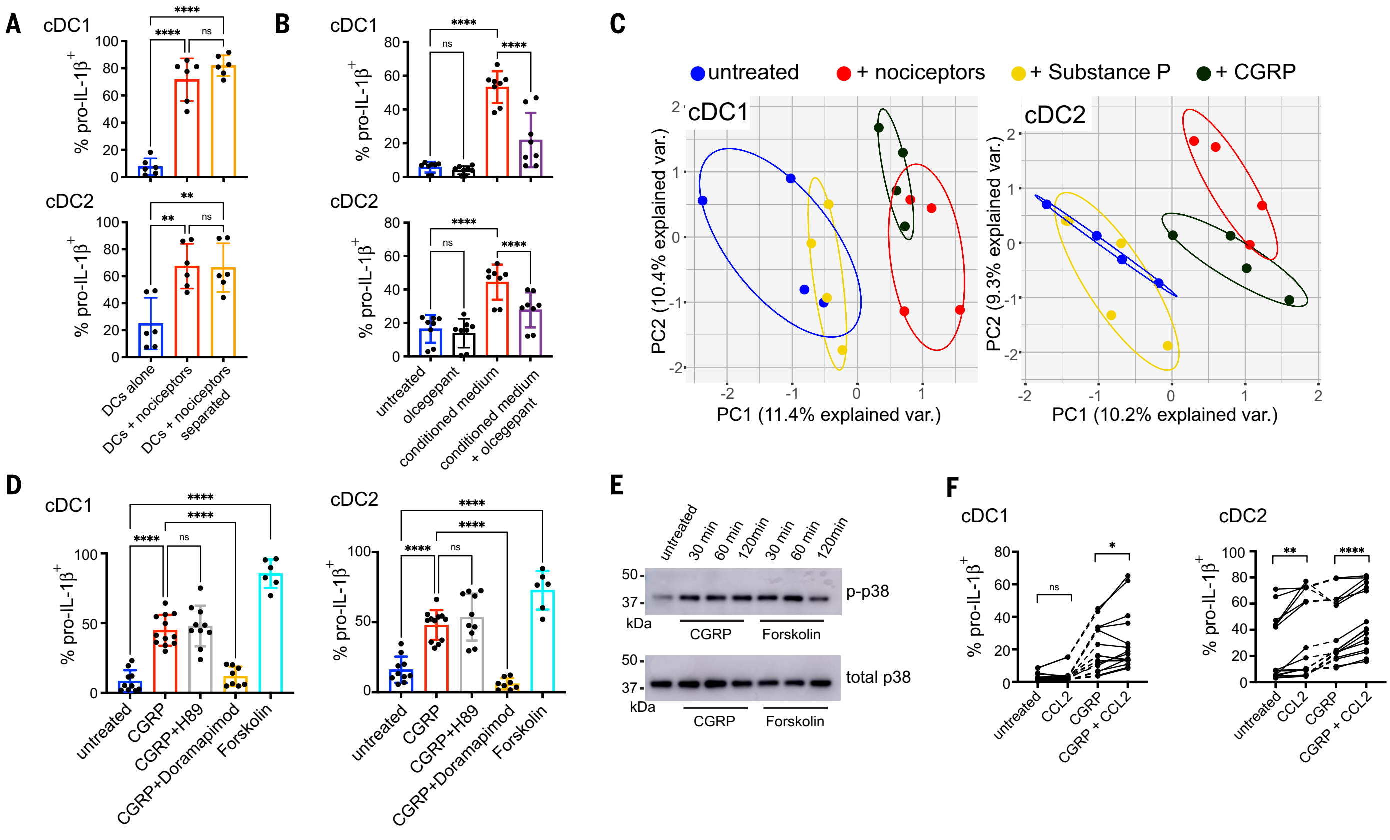


Fig. 6. CGRP induces an enhanced sentinel phenotype in DCs. (**A** and **B**) The intracellular content of pro–IL-1β was determined by flow cytometry in DCs that were directly cocultured with nociceptors, shared medium with nociceptors, or cultured alone (A) and in DCs that were incubated with nociceptor-conditioned medium in the presence or absence of CGRP receptor antagonist (olcegepant) (B). Summary of three to four independent experiments (n = 6 to 8) is shown. (**C**) PCA of RNA-seq data from DCs cultured alone, cocultured with nociceptors, or treated with CGRP or substance P. (**D**) DCs were treated with CGRP, CGRP + dorapamimod (p38 inhibitor), CGRP + H89 (PKA inhibitor), or forskolin (AC activator) overnight. The expression of pro–IL-1β was determined by flow cytometry. Summary of three to six experiments per condition is shown (n = 6 to 12). (**E**) DCs were treated with CGRP or forskolin for the indicated periods of time and lysed. Phosphorylation status of p38 was assessed by phospho-Thr180/Tyr182–specific antibody in immunoblot analysis (top). Total p38 was used as a loading control (bottom). One representative of three independent experiments is shown. (**F**) DCs were plated in wells coated with CCL2 and treated with CGRP overnight. Pro–IL-1β up-regulation was assessed by flow cytometry. Summary of eight independent experiments (n = 16) is shown. Across all panels, data represent mean ± SD. One-way ANOVA with Tukey's multiple-comparisons test (A), two-way ANOVA with Tukey's multiple-comparisons test [(B) and (D)], or paired t test (F) were used for statistical analysis: $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

subset-specific effects of nociceptors likely reflect the fact that cDC1s and cDC2s perform distinct roles in vivo. Nonetheless, the genes co-regulated by nociceptors in cDCs included a number of genes coding for proteins involved in cDC sentinel functions, including pro–IL-1β (fig. S9D and table S1).

CGRP induces an enhanced sentinel phenotype in DCs

Unlike many other proinflammatory cytokines, IL-1β is synthesized as a pro-protein (pro–IL-1β), which is proteolytically processed and released upon inflammasome activation (*41*). Consistent with the lack of overt DC activation in unstimulated cocultures, mature IL-1β was undetectable in coculture supernatants (fig. S10A). However, nociceptors induced a pronounced, contact-independent intracellular accumulation of pro–IL-1β in DCs (Fig. 6A), which was reproduced by exposure to nociceptor-conditioned medium (fig. S10B), suggesting the involvement of a soluble signal(s) such as a neuropeptide(s). Indeed, although SP, PACAP, and VIP were inactive, adrenomedullin and intermedin induced pro–IL-1β expression at submicromolar concentrations, and αCGRP was singularly potent even at picomolar concentrations (fig. S10C). Accordingly, CGRP was readily detectable in nociceptor-conditioned medium and further increased after nociceptor stimulation with capsaicin (fig. S10D). The addition of a CGRP receptor antagonist, olcegepant, to nociceptor-conditioned medium attenuated this effect (Fig. 6B). CGRP up-regulated pro–IL-1β in DCs within 1 hour, reaching a plateau within 4 hours (fig. S10E). Therefore, CGRP pretreatment enhanced the release of mature IL-1β from DCs upon inflammasome activation (fig. S10F). Finally, CGRP admixture to IMQ also significantly increased pro–IL-1β up-regulation in activated DCs. Simultaneously, CGRP at high concentrations exerted a moderate inhibitory effect on IL-12 p40 up-regulation (fig. S10G). This cytokine pattern implies that the DC response to CGRP is independent of and distinct from canonical DC activation mediated by PAMP receptor ligation.

To further explore the impact of nociceptor-derived CGRP, we compared the transcriptomes of DCs cocultured with nociceptors with those treated with CGRP or SP. Exposure to SP did not significantly alter the transcriptional profile of DCs (Fig. 6C and fig. S11A), in agreement with the present data and with published datasets (*10*, *42*), which indicates a lack of SP receptors on BMDCs (fig. S11B). Conversely, CGRP treatment resulted in profound transcriptional changes, reminiscent of, but not identical to, those observed in nociceptor cocultures

(Fig. 6C and fig. S11, A and C), suggesting that CGRP is a major, but not the only, nociceptor-derived signal modulating steady-state DCs. The accumulation of CGRP in the culture media of unstimulated neuronal cultures (fig. S10D) explains why the transcriptional profile of DC–nociceptor cocultures with or without capsaicin showed only minimal differences (fig. S8B). Because of the high sensitivity of DCs toward CGRP, the basal CGRP concentration in conditioned medium was apparently sufficient for a full-fledged DC response, rendering a further capsaicin-induced increase in CGRP functionally irrelevant.

CGRP enhances pro–IL-1β in DCs by signaling through a cAMP-p38–dependent pathway

The CGRP receptor comprises a G protein–coupled receptor (GPCR), which is the signal-transducing unit, and receptor-activity modifying protein 1 (RAMP1), which confers ligand specificity (*43*). Several G-protein subunits interact with this receptor, including $\alpha_{i/o}$, α_q, and α_s. Functionally, $G\alpha_{i/o}$ signaling reduces adenylyl cyclase (AC) activity, whereas $G\alpha_s$ signaling activates AC and promotes cyclic AMP (cAMP) accumulation, activating protein kinase A (PKA) (*43*) and certain mitogen-activated protein kinases (MAPKs) (*44*). To determine the pathways involved in CGRP-mediated pro–IL-1β up-regulation in DCs, we tested small-molecule inhibitors and activators of the respective enzymes. A PKA inhibitor, H89, showed no effect, whereas a p38 MAPK inhibitor, doramapimod, abrogated CGRP-induced pro–IL-1β expression. Conversely, an AC activator, forskolin, triggered pro–IL-1β expression in DCs, even in the absence of CGRP (Fig. 6D). Both CGRP and forskolin induced rapid phosphorylation of p38 (Fig. 6E), consistent with a model in which p38 activation lies downstream of CGRP-mediated AC activation. Additionally, 8-Br-cAMP, a hydrolysis-resistant synthetic analog of cAMP, drove pro–IL-1β accumulation in DCs (fig. S11D), and pertussis toxin–mediated inhibition of $G\alpha_{i/o}$ signaling further potentiated the effect of CGRP (fig. S11E).

Finally, surface-immobilized CCL2 enhanced the DC response to CGRP (Fig. 6E). CCR2 is a $G\alpha_i$-coupled GPCR, so CCR2 signaling might limit the availability of $G\alpha_i$ subunits, resulting in enhanced $G\alpha_s$ coupling to the CGRP receptor and increased cAMP signaling. Thus, in DCs, CGRP receptor signaling through $G\alpha_s$ promotes cAMP accumulation and initiates a p38-dependent program of gene expression (fig. S11F). Indeed, GSEA confirmed an enrichment of MAPK and p38 MAPK pathway genes in DCs after coculture with nociceptors (fig. S11G). This reprogramming is molecularly distinct from classical activation and renders DCs in a poised state of enhanced sentinel function characterized by intracellular accumulation of pro–IL-1β, as well as the up-regulation of genes important for pathogen resistance, regulation of phagocytosis, cell adhesion and migration, cytokine responsiveness, and antigen presentation (table S1).

Nociceptors promote cytokine production by dermal DCs in vivo

Having identified three distinct mechanisms by which nociceptors modulate DC functions in vitro, electrical activity, release of CGRP, and secretion of CCL2, we sought to determine their in vivo relevance by examining mice in which nociceptors had been ablated by treatment with resiniferatoxin (RTX). As shown previously (*9*) in mice with intact nociceptors, topical treatment with IMQ stimulated the production of IL-12 p40 and an influx of monocytes and neutrophils, a response that was markedly attenuated in RTX-treated animals (Fig. 7A and fig. S12A). Consistent with our in vitro observations, this correlated with a markedly lower frequency of IL-12 p40$^+$ dermal DCs in nociceptor-depleted mice (fig. S12B) even though the overall number of dermal DCs in steady-state skin was unaffected (fig. S12C). Similarly, blocking nociceptor electrical activity by topical treatment with lidocaine and QX-314 mimicked the effect of denervation and prevented IMQ-induced IL-12 p40 accumulation, as well as monocyte and neutrophil influx into the skin (Fig. 7B and fig. S12D). Finally, as predicted in vitro, both nociceptor depletion and anesthetics also decreased the accumulation of IL-6 (Fig. 7, C and D). Thus, the ablation of nociceptors and the pharmacological silencing of their electrical activity using topical anesthetics ameliorate the pathology that drives DC-dependent psoriasiform skin inflammation.

CGRP induces pro–IL-1β up-regulation in dermal DCs in vivo

In contrast to our in vitro experiments, in which the presence of nociceptors was sufficient to up-regulate pro–IL-1β in DCs (Fig. 6A), the in vivo ablation of nociceptors did not alter baseline levels of pro–IL-1β in dermal DCs (Fig. 7E). This was likely due to the fact that nociceptors in vitro spontaneously release CGRP (*45*), whereas resting nociceptors in vivo do not (*46*). Accordingly, when control mice were topically challenged with IMQ, dermal DCs responded by upregulating pro–IL-1β (Fig. 7E), resulting in the accumulation of mature IL-1β in the tissue (Fig. 7F). By contrast, IMQ had only a minimal effect on pro–IL-1β expression and IL-1β accumulation in animals in which nociceptors had been depleted (Fig. 7, E and F) or silenced (Fig. 7, G and H) in a manner analogous to the IL-12 p40 response. However, local treatment with capsaicin, a stimulus for nociceptors, but not DCs, failed to induce IL-12 p40 (fig. S12E), whereas pro–IL-1β was readily up-regulated in nociceptor-sufficient animals, an effect that was significantly dampened in RTX-treated mice (Fig. 7I). These findings were consistent with our in vitro observations that neuronal activation alone was insufficient to drive IL-12 p40 expression in otherwise unstimulated DCs (Fig. 1I). Capsaicin treatment also weakly induced pro–IL-1β up-regulation in RTX-treated animals (Fig. 7I), possibly because of incomplete nociceptor depletion or capsaicin-mediated activation of cutaneous mast cells, which may express TRPV1 (*47*), CGRP (fig. S12F), and other factors that could induce pro–IL-1β up-regulation in DCs (*48*).

To determine whether the induction of pro–IL-1β in DCs in vivo is CGRP dependent, we induced sterile tissue inflammation by intradermal corn oil injection. Admixture of olcegepant to the injectate decreased pro–IL-1β up-regulation in DCs, but had no impact on the influx of inflammatory cells (Fig. 7J and fig. S12G). Thus, the activation of nociceptors and resultant local release of CGRP are sufficient to drive pro–IL-1β up-regulation in dermal DCs independently of other inflammatory signals.

Nociceptor-derived CCL2 controls DC-dependent dermal immune responses

Finally, we assessed the role of nociceptor-derived CCL2. First, to determine CCL2 expression in nociceptors, we examined *Ccl2-mCherry*$^{fl/fl}$ reporter mice, which express a CCL2-mCherry fusion protein deletable by Cre recombinase (*49*). CCL2-mCherry was readily detectable by confocal microscopy within intra-axonal vesicle–like structures in cutaneous fibers at steady state (Fig. 8A) and after IMQ treatment (fig. S13A). CCL2 expression was not restricted to nociceptors, indicating that nociceptors are not the sole source of CCL2 in murine skin. Indeed, when *Ccl2-mCherry*$^{fl/fl}$ mice were crossed with *Scn10a*$^{Cre/+}$ mice (hereafter called NaV1.8$^{\Delta CCL2}$), the CCL2 reporter signal in non-neuronal cells remained unchanged but was extinguished in cutaneous nerve fibers (Fig. 8B and fig. S13B).

Upon topical treatment with IMQ, NaV1.8$^{\Delta CCL2}$ animals exhibited smaller inflammatory infiltrates (Fig. 8C) and decreased IL-12 p40 accumulation in the skin (Fig. 8D) compared with Cre-negative littermates. Dermal DCs expressed high levels of CCR2 (fig. S13C), indicating that they were equipped to sense and respond to CCL2. Nonetheless, the number of DCs in unperturbed skin was comparable between NaV1.8$^{\Delta CCL2}$ animals and littermate controls (fig. S13D), suggesting that nociceptor-derived CCL2 was dispensable for the maintenance of dermal DCs under steady-state conditions. After IMQ treatment, however, NaV1.8$^{\Delta CCL2}$ mice showed a decrease in dermal DC numbers compared with their littermates (Fig. 8E), suggesting that nociceptor-derived CCL2 serves as a retention signal preventing DCs from prematurely leaving the skin, thus allowing for a

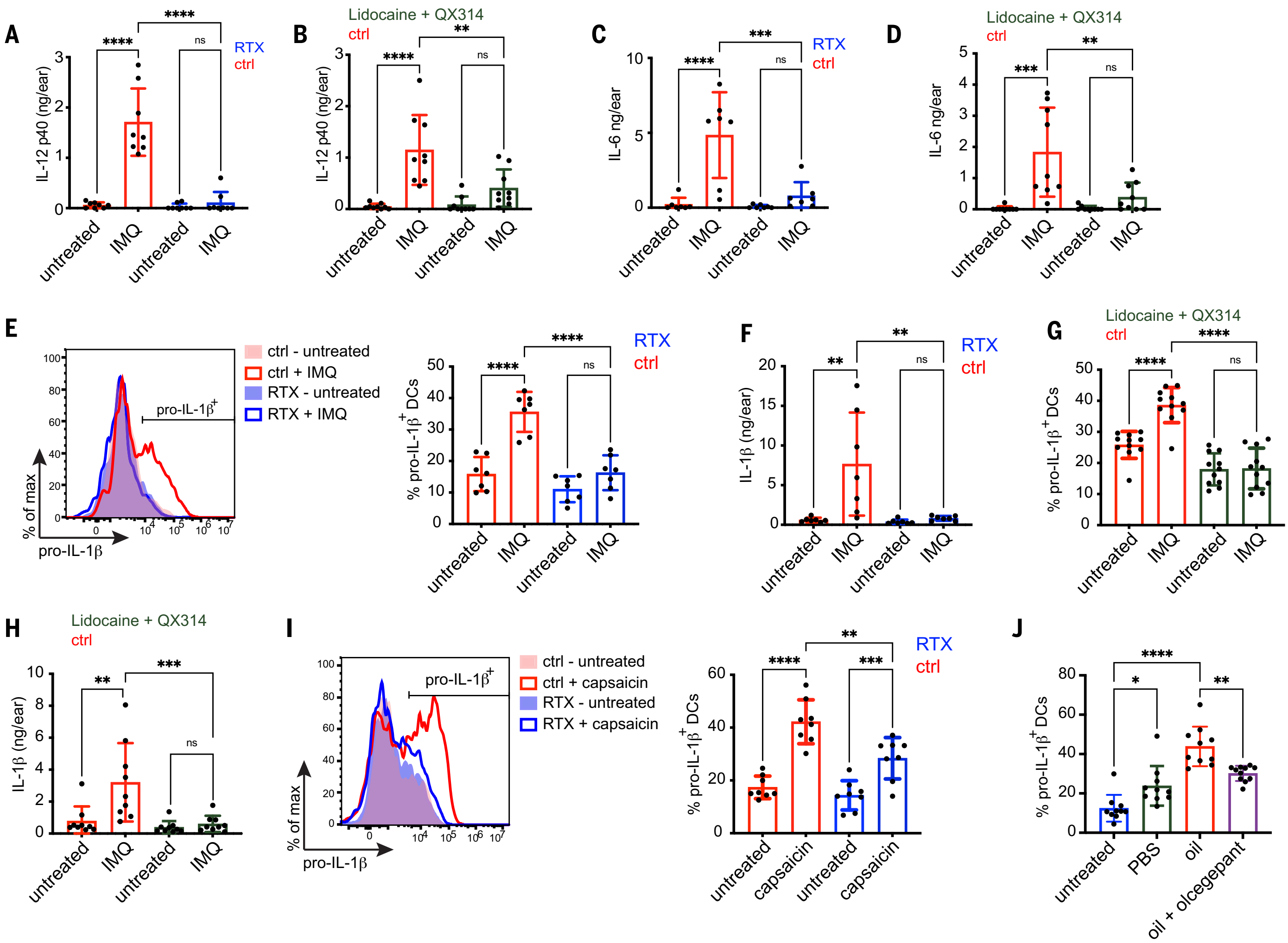


Fig. 7. Nociceptors control DC functions in murine skin in vivo. (**A** to **D**) Ears of mice in which nociceptors had been left intact, were chemically ablated, or were treated with lidocaine + QX314 were then treated topically with IMQ cream. Cytokine accumulation was analyzed by tissue lysate ELISA. Summary of three to four experiments (n = 7-9 per group) is shown. (**E**) Ears of mice in which nociceptors had been left intact or were chemically ablated were treated with IMQ cream, and pro–IL-1β expression by dermal DCs was assessed by flow cytometry. One representative experiment (left) and quantification of three independent experiments (n = 7 per group; right) are shown. (**F**) Ears of mice were treated as in (E) and the accumulation of mature IL-1β cytokine in the tissues was assessed by by tissue lysate ELISA. Summary of three experiments (n = 7 per group) is shown. (**G**) Ears of mice in which nociceptors had been left intact or ears treated with lidocaine + QX314 were treated topically with IMQ cream, and pro–IL-1β expression by dermal DCs was assessed by flow cytometry. Summary of five experiments (n = 11 per group) is shown. (**H**) Ears of mice were treated as in (G) and the accumulation of mature IL-1β cytokine in the tissues was assessed by by tissue lysate ELISA. Summary of three experiments (n = 9 per group) is shown. (**I**) Ears of mice in which nociceptors had been left intact or were chemically ablated were treated with capsaicin and analyzed for pro–IL-1β expression by flow cytometry. One representative experiment (left) and quantification of four independent experiments (n = 8 per group; right) are shown. (**J**) Mouse ears were injected intradermally with PBS, oil, or oil supplemented with olcegepant, and pro–IL-1β up-regulation in DCs was analyzed by flow cytometry 16 hours later. Quantitation of four independent experiments (n = 10 to 11 per group) is shown. Across all panels, data represent mean ± SD. Two-way ANOVA with Tukey's multiple-comparisons test was used for statistical analysis: *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

longer period of in situ cytokine production and more pronounced local inflammation.

In addition to compromising local inflammatory responses, we reasoned that the dysregulation of DC migration may also interfere with the acquisition of cutaneous antigens and affect the initiation of adaptive immune responses. To test this idea, we performed contact hypersensitivity (CHS) experiments in which the adaptive recall response depends on the acquisition of hapten-modified self-antigens by dermal DCs and their presentation to adaptive lymphocytes in draining lymph nodes (*50*). In naïve mice, exposure of ear skin to hapten caused a comparable irritant response in NaV1.8$^{\Delta CCL2}$ animals and littermates. By contrast, the DC-dependent recall response in mice sensitized by prior hapten painting of abdominal skin was nearly abrogated in the absence of nociceptor-derived CCL2 (Fig. 8F and fig. S13E). Nociceptor ablation has previously been shown to impair the ability of DCs to acquire viral antigens in HSV-1–infected skin (*51*), suggesting that this phenotype of NaV1.8$^{\Delta CCL2}$ mice was likely caused by a defect in DC-dependent priming of antigen-specific lymphocytes in lymph nodes. Indeed, when WT mice were treated with lidocaine and QX-314 to transiently inhibit nociceptor activity in abdominal skin during sensitization, the CHS response was compromised

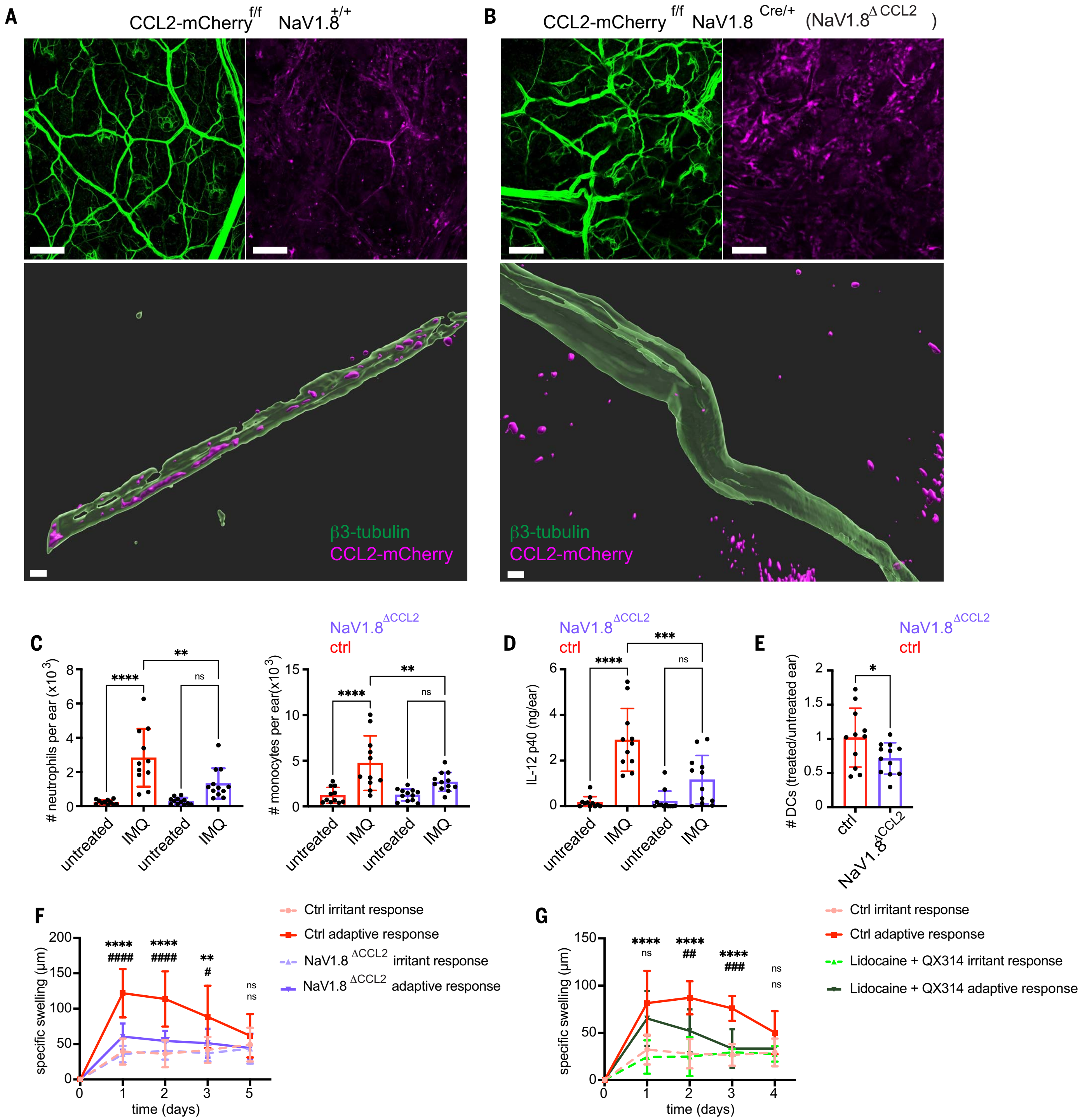


Fig. 8. Nociceptor-derived CCL2 is important for induction of local inflammatory and adaptive immune responses. (**A** and **B**) Dermal sheets from NaV1.8$^{+/+}$ CCL2-mCherry (control) (A) and NaV1.8$^{Cre/+}$ CCL2-mCherry (NaV1.8$^{\Delta CCL2}$) (B) animals were fixed, stained with anti–β3-tubulin (green) and anti-mCherry (magenta) antibodies, and analyzed using confocal microscopy. Examples of neuronal fibers showing CCL2-mCherry signal are highlighted with arrows (top panels). Surface reconstructions of high-resolution images were performed in Imaris to illustrate the vesicle-like staining pattern of CCL2 (bottom panels). Scale bars: 100 μm in top panels and 2 μm in bottom panels. (**C** to **E**) Ears of NaV1.8$^{\Delta CCL2}$ or littermate control mice were treated with IMQ cream and inflammatory infiltrate (C) and IL12-p40 accumulation (D) and ratios of DCs between treated and untreated ears for each animal (E) were analyzed. Summary data of six experiments (n = 11 to 12 per group) is shown. (**F** and **G**) Mice were sensitized with DNFB or treated with vehicle control (acetone) and challenged 5 days later with DNFB on one ear and acetone on the other. Ear thickness was measured at the indicated time points after challenge, and specific swelling was calculated as the difference between the DNFB-treated ear and the vehicle-treated ear for each animal. (F) NaV1.8$^{\Delta CCL2}$ or littermate control mice (n = 7 to 8 per group). (G) WT mice treated with lidocaine + OX314 before

 https://avxhm.se/blogs/hill0

and on 2 consecutive days after sensitization. All data are presented as mean ± SD. Two-way ANOVA with Tukey's multiple-comparisons test was used for statistical analysis. For (F) and (G), *statistical comparison between control DNFB-sensitized and control nonsensitized groups and #statistical comparison between control DNFB-sensitized and NaV1.8$^{\Delta CCL2}$ DNFB-sensitized groups (F) or between control DNFB-sensitized and lidocaine + QX314 DNFB-sensitized groups (G). *, #$P < 0.05$; **, ##$P < 0.01$; ***, ###$P < 0.001$; ****, ####$P < 0.001$.

in subsequently challenged ear skin where nociceptors had always been fully functional (Fig. 8G and fig. S13F). Thus, nociceptors must be functional and express CCL2 for the proper induction of both DC-dependent local inflammation and the priming of adaptive immune responses against dermal antigens.

Discussion

At the transcriptome level, our results indicate that the presence of nociceptors induces numerous subset-specific changes in cDC1 and cDC2 cells in vitro. These differential responses likely reflect the fact that cDC1s and cDC2s have distinct roles in regulating immune responses, and it will be important to determine whether and how nociceptors regulate subset-specific cDC functions in vivo. The present study has, however, focused on exploring the communication signals emanating from nociceptors that elicit a shared response by all cDCs in peripheral tissues.

We show here that nociceptors control DC functions in a context-dependent manner through at least three independent communication modalities. Without immune stimuli, nociceptor-derived CGRP induces transcriptional changes in steady-state DCs to up-regulate sentinel function–related genes but without causing overt DC activation. Upon concomitant encounter of painful stimuli and PAMPs, the nociceptors' electrical activity is sensed by interacting PAMP-stimulated DCs, which increases the proportion of DCs that commence the production of proinflammatory cytokines. Thus, DCs are uniquely equipped to sense and interpret nociceptor activation either as a "reminder" to assume a poised, anticipatory state when other inflammatory signals are absent or as an "alarm call" that amplifies their collective inflammatory response to immune stimuli. Additionally, by secreting CCL2, nociceptors further fine-tune DC functions by promoting their contribution to local inflammatory responses and by regulating the egress of DCs from the skin to optimize adaptive immune responses (fig. S14).

CGRP receptor–mediated activation of AC (*52*) exerts anti-inflammatory effects on immune cells, including macrophages (*53*), neutrophils (*46*), γδT cells (*54*), and innate lymphoid cells (*55*). By contrast, CGRP, even at high concentrations, had only modest inhibitory effects on IL-12 p40 up-regulation in activated BMDCs (figs. S4E and S10G). This was in contrast to previously reported inhibitory effects of CGRP on IL-12 production in granulocyte-macrophage colony-stimulating factor (GM-CSF)–induced BMDCs (*11*). Unlike the Flt3L-induced BMDCs used here, GM-CSF cultures contain macrophages (*56*), which are potently inhibited by nociceptors (Fig. 1D) and by CGRP (*52*). It is therefore possible that the negative effect of CGRP on GM-CSF–induced BMDC cultures may be a consequence of macrophage contamination rather than a true effect on DCs. Finally, taken at face value, these observations also suggest an inherent difference in the intracellular signaling pathways between DCs and macrophages whereby AC activation may exert strong inhibitory effects on the latter (*57*) but not the former.

In light of the exquisite sensitivity of DCs toward CGRP, the close proximity of DCs and nociceptors in vivo, and the anticipatory nature of the DC response elicited by CGRP, it is tempting to speculate that DCs may have specifically evolved to register and interpret a local, limited release of CGRP induced by transient or subtle activation of nociceptors as a warning signal that could portend a barrier breach. By contrast, other immune cells that do not a priori associate closely with nociceptors may only sense CGRP when it is released in larger quantities, likely as a consequence of severe tissue damage. In this context, CGRP-induced anti-inflammatory and tissue repair–promoting functions [e.g., through action on macrophages (*57*) or neutrophils (*46*)] may become more prominent.

Although CGRP did not induce IL-23 production by DCs in this study, recent work has demonstrated that nociceptor-derived CGRP is sufficient and necessary for IL-23 production by CD301b$^+$ dermal DCs during *C. albicans* infection (*8*). The reasons for this discrepancy remain unclear, but it is possible that CGRP-induced IL-1β potentiates IL-23 expression (*58*, *59*) or that an intermediary cell type(s) that generates as-yet-unidentified secondary signals in response to CGRP to promote IL-23 production by DCs may be involved.

Aside from neuropeptides, our results highlight a critical role for CCL2, a chemokine whose expression by nociceptors has traditionally been considered a sign of pathology. CCL2 can recruit inflammatory cells to DRGs, resulting in peripheral sensitization and neuropathic pain (*60*). However, some nociceptors express CCL2 even at steady state, at least at the mRNA level (*3*, *61*, *62*). Indeed, we show here that the expression of CCL2 in cutaneous fibers is not necessarily maladaptive but rather serves to fine-tune DC-dependent immune responses. Aside from DCs, other cell types, including monocytes and memory T-cell subsets, also express the CCL2 receptor CCR2; however, the impact of nociceptor-derived CCL2 on these populations remains to be explored. Moreover, further work will be needed to determine how CCL2 is regulated in nociceptors. In vitro, cultured nociceptors released CCL2 spontaneously, and nociceptor stimulation further potentiated CCL2 release, similar to the dynamics of neuropeptide release. In vivo, neuropeptide release from nociceptors is tightly controlled (*46*), but it remains to be established whether similar rules apply to CCL2.

Although myeloid leukocytes are generally not considered excitable, we observed that activation of nociceptors induced membrane depolarization and calcium flux in the interacting DCs. Spreading of action potentials to and among non-neural cells through mechanisms including gap junctions and tunneling nanotubes has been described previously (*63*), and both of these communication modalities have been reported for DCs (*64*). Whatever the mechanism, our observations establish a precedent for a direct, contact-dependent, and neuropeptide-independent communication pathway between nociceptors and DCs, which culminates in calcium mobilization in the latter. Intracellular calcium is a key second messenger in numerous signaling processes. The downstream molecular signaling pathway(s) promoting DC production of IL-12 p40 and IL-6 but not TNFα remains to be clarified. Nonetheless, synergy between MyD88- and TRIF-dependent pathways in DCs has previously been shown to stimulate IL-12 and IL-6 secretion while having no effect on TNFα (*65*), indicating that these cytokines are not coregulated. Conceivably, calcium mobilization in activated DCs triggered by firing nociceptors might act in an analogous manner to synergize with TLR-induced MyD88 signaling to potentiate the expression of IL-12 p40 and IL-6.

Materials and Methods

Mice

C57BL/6J (JAX stock no. 000664), *Ccl2*$^{-/-}$ (JAX stock no. 004434) (*26*), *Tlr7*$^{-/-}$ (JAX stock no. 008380) (*17*), *Zbtb46*^{eGFP} (JAX stock no. 027618) (*30*), and *Ccl2-mCherry*$^{fl/fl}$ (JAX stock no. 016849) (69) mice were purchased from The Jackson Laboratory. *I-A*^{b}β-eGFP (*66*), *Scn10a*^{Cre} (*67*), and *Scn10a*$^{Cre/Cre}$*Rosa26*$^{TdT/TdT}$ (*3*) mice were all described previously and were bred at the Harvard Medical School animal facility. All animal experiments were performed in accordance with national and institutional

guidelines, and were approved by the institutional animal care and use committee and the Committee on Microbiological Safety (COMS) of Harvard Medical School. Both male and female mice were used.

In vitro DC generation

Flt3L-induced BM-derived dendritic cells were generated as described previously (68). Briefly, BM was harvested from femurs and tibiae of 6- to 8-week-old C57BL/6J mice of either sex. Red blood cells were lysed and BM cells were plated with 100 ng/ml recombinant Flt3L (R&D Systems) in six-well plates (3 ml per well), one plate per mouse in RPMI1640 medium (Gibco) containing 10% fetal bovine serum (FBS) (Gemini Bio), β-mercaptoethanol, glutamine, and penicillin–streptomycin. Fully developed DCs were harvested and used on days 9 and 10.

In vitro macrophage generation

M-CSF–induced BM macrophages were generated as described previously (69). Briefly, BM was harvested from femurs and tibiae of 6- to 8-week-old C57BL/6J mice of either sex. Red blood cells were lysed and BM cells were plated with 25 ng/ml recombinant M-CSF in six-well plates (3 ml per well) in RPMI1640 medium (Gibco) containing 10% FBS (Gemini Bio), β-mercaptoethanol, glutamine, and penicillin–streptomycin. Fully developed macrophages were harvested and used on day 7.

In vitro nociceptor culture

DRG cultures were prepared from freshly isolated dorsal root ganglia from adult (6- to 10-week-old) C57BL/6J mice of either sex as described previously (45, 70) with minor modifications. Briefly, the dissection was performed in ice-cold phosphate-buffered saline (PBS), and harvested DRGs were kept on ice throughout. The first step of digestion was performed for 10 min at 37°C in 3 ml of Hank's balanced salt solution (HBSS) (Gibco) with 60 U of papain (Worthington), 0.5 mM EDTA, and 1.5 mM $CaCl_2$. The second digestion step was performed in 3 ml of HBSS with 4 mg/ml collagenase Type-2 (Worthington) and 5 mg/ml dispase (Gibco) for 30 min at 37°C. Cell suspensions were triturated by repeated pipetting in Leibovitz medium (Gibco) supplemented with 10% FBS (Gemini Bio), overlaid over 20% Percoll (GE Healthcare) in Leibovitz medium, and centrifuged for 9 min at 400*g* with no brake. Pelleted cells were washed in Neurobasal Medium (Gibco) containing B27 supplement (Gibco) and penicillin–streptomycin before being plated (1-1.5×10^4 cells per well) in wells precoated with poly-D lysine and laminin (both Sigma Aldrich) in Neurobasal Medium containing B27 supplement and penicillin–streptomycin, supplemented with 5 μM cytosine β-D-arabinofuranoside (Sigma-Aldrich) and 25 ng/ml mouse recombinant NGF (R&D Systems). Culture medium was refreshed every 4 to 5 days, and neuronal culture was typically used between days 7 and 10 after plating. Cultures were visually inspected to verify neuronal recovery and depletion of contaminating glial/Schwann cells before each experiment.

In vitro cortical and hippocampal neuron cultures

CNS neuronal cultures were prepared from freshly isolated C57BL/6J mouse embryonic day 18 (E18) to E19 brains, as described previously (71) with minor adjustments. Dissection was performed in ice-cold HBSS + HEPES (both Gibco), and tissues were digested at 37°C for 15 min in 0.25% trypsin (Invitrogen) for hippocampus and 0.25% trypsin + 0.5 mg/ml DNAse I (Roche) for cortex samples. Tissues were homogenized by trituration in a fire-polished Pasteur pipette, and single-cell suspensions were generated by filtration through 100-μm cell strainers. Cells were washed in Neurobasal Medium containing B27 supplement (both Gibco) and penicillin–streptomycin before being plated in wells precoated with poly-D-lysine and laminin (both Sigma-Aldrich) in Neurobasal Medium containing B27 supplement and penicillin–streptomycin supplemented with 5 μM cytosine β-D-arabinofuranoside (Sigma-Aldrich). Culture medium was refreshed every 3 days, and the culture was used between days 7 and 14 after plating.

In vitro cocultures

For coculture experiments, neuronal medium was replaced with fresh Neurobasal Medium containing B27 supplement and penicillin–streptomycin, and 1 × 10^5 BMDCs per well were added in an equal volume of RPMI 1640 medium containing 10% FBS, β-mercaptoethanol, glutamine, and penicillin–streptomycin. Cells were allowed to interact for 2 to 4 hours before any treatments or stimuli were applied unless stated otherwise. As a control, DCs were plated and treated in an identical fashion in wells containing an appropriate volume of Neurobasal Medium with B27 supplement and penicillin–streptomycin but no nociceptors.

In vitro treatments

All treatments were performed overnight unless otherwise stated. All CGRP treatments were performed with 1 nM CGRP unless otherwise indicated. Lidocaine (500 μM), QX314 (1 mM), A967079 (10 μM), and olcegepant (100 nM) were added to cultured nociceptors 2 hours before the addition of DCs. H89 (1 μM), doramapimod (1 μM), and pertussis toxin (100 ng/ml) were added to DCs immediately before CGRP treatment. 8-Br-cAMP (1 mM) and forskolin (50 μM) were used without further stimuli. A23187 (400 ng/ml) was added to DCs alongside IMQ, and the treatment lasted for 4 to 5 hours to avoid toxicity. Activating stimuli were used at the following concentrations: IMQ, 1 μg/ml; zymosan, 10 μg/ml; LPS, 1 ng/ml; flagellin, 20 ng/ml; CpG, 0.1 μM; poly I:C, 1 μg/ml; PR-8 GFP IAV, multiplicity of infection (MOI) = 0.25; *S. pneumoniae*, MOI = 5; and *C. albicans*, MOI = 10.

Microbes

Influenza PR8-GFP was generously provided by Dr. A. Garcia-Sastre at the Mount Sinai School of Medicine and the NIAID Centers of Excellence for Influenza Research and Surveillance program. *S. pneumoniae* was a kind gift from Dr. R. Malley at the Boston Children's Hospital and Harvard Medical School. *C. albicans* was purchased from ATCC (MYA-2876).

ELISA

ELISA kits specific for IL-12 p40 (catalog no. 431601), IL-6 (catalog no. 431302), TNFα (catalog no. 430902), IFN-β (catalog no. 439407), CCL2 (catalog no. 432704), and IL-1β (catalog no. 432603) were from BioLegend. The EIA kit specific for CGRP (catalog no. 589001) was from Cayman Chemicals. All assays were performed as per the manufacturers' instructions.

In vitro coculture flow cytometry analysis

In vitro cultures were harvested in warm 5 mM EDTA in PBS. Dead cells were then stained using the LIVE/DEAD Fixable Near-IR Dead Cell Stain Kit (Thermo Fisher) in ice-cold PBS. Samples were washed and stained with fluorescent antibodies for appropriate surface markers in fluorescence-activated cell sorting (FACS) buffer (1% FBS, 5 mM EDTA, and 0.1% NaN_3 in PBS) on ice. Intracellular staining was performed using the Cytofix/Cytoperm kit (BD Biosciences) according to the manufacturer's instructions. cDC1s were identified as XCR-1$^+$, and cDC2s were identified as CD11b$^+$. For IL-6 intracellular stains, 10 μM brefeldin A was added to the cultures 2 hours before harvest.

Transwell assays

BMDCs (0.5 to 1 × 10^6) in complete RPMI 1640 medium were added to the upper compartment of a Transwell setup (Corning) with 5-μm pores and allowed to migrate into the lower compartment containing chemotactic stimuli for 4 hours. Migrated cells were harvested, stained with fluorescent antibodies, and enumerated by flow cytometry. The chemotactic index was calculated as the number of cells transmigrated under a given condition divided by the number of cells that transmigrated between two compartments containing the same medium.

Chemotactic molecule identification

The Mouse Chemokine Array C1 (Ray Biotech) and Proteome Profiler Mouse Chemokine Array Kit (R&D) were used. Membranes were incubated with nociceptor-conditioned medium or control medium overnight and developed

as per the manufacturers' instructions. The signal intensity for each spot was quantified in Fiji 2.0.0 (72) using the Analyze Gels function.

SDS-PAGE and immunoblot analysis

Cells were harvested and lysed for 30 min in Tris buffer, pH 7.5, with 1% SDS, 150 mM NaCl, 10 mM NaF, 2 mM Na_3VO_4, protease inhibitors (Thermo Scientific), and benzonase (Sigma Aldrich). After 30 min, the lysate was diluted 1 to 10 into Tris buffer, pH 7.5, with 1% NP40, 150 mM NaCl, 10 mM NaF, 2 mM Na_3VO_4, and protease inhibitors (Thermo Scientific) and centrifuged for 15 min at 16,000*g*. Supernatants were prepared for SDS–polyacrylamide gel electrophoresis (SDS-PAGE) in 6× Laemmli buffer containing 60% v/v glycerol, 150 mg/ml SDS, and 0.75 mg/ml bromophenol blue in 75 mM Tris–HCl, pH 6.8, in the presence of 100 mM dithiothreitol. Separation was performed using 4 to 12% NuPAGE precast gels (Invitrogen). Proteins were transferred onto Immobilon P membrane (Merck Millipore) using wet transfer, and the membrane was blocked in 5% bovine serum albumin (BSA) in PBS + 0.05% Tween-20. Binding of appropriate horseradish peroxidase–conjugated antibodies was revealed using Luminata Forte HRP substrate (Merck Millipore) and Amersham Imager 600 (GE Healthcare).

Live-cell imaging

Live-cell imaging was performed on a Nikon Ti inverted microscope equipped with W1 Yokogawa Spinning disk with 50-µm pinholes, an Andor Zyla 4.2 Plus sCMOS monochrome camera, and the OKO Lab Heated enclosure with environmental control set to 37°C and 5% CO_2. A Plan Apo λ 20X/0.75 DIC I objective was used. All imaging experiments were performed in a 1:1 mixture of RPMI 1640 medium containing 10% FBS, β-mercaptoethanol, glutamine, and penicillin–streptomycin and Neurobasal Medium containing B27 supplement and penicillin–streptomycin, both phenol red–free, on µ-Slide eight-well chambered coverslips (Ibidi). For the microfluidics experiments, Xona Microfluidics chips with 150-µm microgrooves were used.

For calcium-imaging experiments, cells were loaded with 10 µM Fluo-4 calcium-sensitive dye (Thermo-Fisher) for 20 min at 37°C. Immediately before imaging, Fluo-4–containing medium was replaced with fresh, prewarmed, phenol red–free medium, and the Fluo-4 signal was imaged in a single confocal plane using a 488-nm laser line and a GFP filter cube (Chroma ET 49002).

For membrane potential imaging experiments using BeRST dye, cells were loaded with 5 µM BeRST and 10 µM Fluo-4 for 20 min at 37°C. Immediately before imaging, dye-containing medium was replaced with fresh, prewarmed phenol red–free medium, and images were acquired in a single confocal plane using 488- and 640-nm laser lines and GFP and Cy5 filter cubes (ET Chroma 49002 and 49006).

For membrane potential imaging experiments using Di-8-ANNEPS, cells were loaded with 5 µM Di-8-ANNEPS for 30 min at 37°C. Immediately before imaging, dye-containing medium was replaced with fresh, prewarmed medium, and images were acquired in a single confocal plane using a 488-nm laser line and mCherry and Cy5 filter cubes (ET Chroma 49008 and 49006).

Light-sheet imaging

The lattice light-sheet microscope was constructed as described previously (73) with a Special Optics 0.65 numerical aperture (NA) excitation objective, a Nikon CFI Apo LWD 25XW 1.1 NA detection objective, an Orca Flash 4.0 v3 sCMOS camera (Hamamatsu), and 488-nm (300 mW Coherent Sapphire) and 560-nm (500 mW, MPB Communications) lasers. The annular mask was set at 0.42 to 0.5 NA, and a square lattice in the dithered mode was produced at the sample. The excitation power (488 nm) was measured at the back focal plane of the excitation objective at ~100 µW. The 25× detection objective was paired with a 500-mm achromat lens for an effective magnification of 63.7×, resulting an image pixel size of 102 nm. The exposure time for each plane was 4 ms, and the stage-scanning step size for the volumetric imaging was 0.5 mm, corresponding to 265 nm along the optical axis after deskewing. Data were deskewed and deconvolved with LLSpy (DOI: 10.5281/zenodo.1059099) and cudaDeconv (source code available at https://github.com/scopetools/cudaDecon) using a PSF measured from a 0.1 µm bead. sCMOS residual charge artifact was corrected using LLSpy (https://llspy.readthedocs.io/en/latest/camera.html).

Electron microscopy

Nociceptors were grown on aclar coverslips precoated with poly-D-lysine and laminin. DCs were allowed to interact with nociceptors for 4 hours before the co-culture was fixed in a mixture of 2.5% glutaraldehyde, 1.25% paraformaldehyde, and 0.03% picric acid in 0.1 M sodium cacodylate buffer, pH 7.4, for 1 hour at room temperature. The cells were washed in 0.1 M sodium cacodylate buffer, pH 7.4, postfixed for 30 min in a 1% OsO_4/1.5% $KFeCN_6$ solution, washed twice in water and once in maleate buffer, and incubated in 1% uranyl acetate in maleate buffer for 30 min. This was followed by two washes in water and subsequent dehydration in grades of alcohol (5 min each at 50, 70, and 95% and twice at 100%). The samples were subsequently embedded in TAAB Epon (Marivac Canada Inc.) and polymerized at 60°C for 48 hours. After polymerization, the aclar was peeled off, and ultrathin sections (~80 nm) were cut on a Reichert Ultracut-S microtome, picked up onto copper grids, and stained with lead citrate. Grids were examined under a TecnaiG^2 Spirit BioTWIN transmission electron microscope, and images were recorded using an AMT 2k CCD camera.

Confocal imaging

For whole-mount imaging of dermal sheets, ears of from Zbtb46^{eGFP}Scn10a$^{Cre/+}$ROSA$^{tdTomato/+}$ animals were split into the dorsal and ventral aspects, fixed in 4% paraformaldehyde (PFA), permeabilized in 0.3% Tween in PBS, and stained with S100β primary antibody overnight in a blocking buffer containing 3% BSA. Samples were washed and stained with anti-rabbit Alexa Fluor 647 and anti-GFP Alexa Fluor 488 antibodies. Ears of *Ccl2-mCherry*$^{fl/fl}$ or Scn10a$^{Cre/+}$*Ccl2-mCherry*$^{fl/fl}$ animals were split into the dorsal and ventral aspects, fixed in 4% PFA, permeabilized in 0.3% Tween in PBS, blocked in 3% BSA, and stained with β3-tubulin and mCherry antibodies for 48 hours. Subsequently, the samples were washed and stained with anti-goat Alexa Fluor 568 and anti-rabbit Alexa Fluor 647 antibodies.

For imaging of fixed nociceptor cultures, nociceptors isolated from NaV1.8$^{Cre/+}$Rosa26$^{tdTomato/+}$ were plated on µ-Slide eight-well chambered coverslips (Ibidi), and cultures were maintained as per the standard protocol. Cells were fixed in 4% PFA, permeabilized in 0.3% Tween in PBS, and stained with β3-tubulin primary antibody overnight in a blocking buffer containing 3% BSA. Samples were then washed and stained with anti-rabbit Alexa Fluor 488 antibody.

Images were acquired on an Olympus IX83 inverted single-point laser scanning confocal microscope with UPlan S Apo 20×/0.75 air or UPlan X Apo 60×/1.42 oil objectives.

Image analysis

Calcium imaging and membrane potential analysis were performed in Fiji 2.0.0 (72). Images are presented using the "Fire" LUT, and quantitation of signal was performed in selected regions of interest across all time points using the "Plot Z-axis profile" function to generate single-cell response graphs. For statistical analysis, cells were manually traced and were considered "responders" if their Fluo-4 signal showed at least 50% increase ($F_{max}/F_0 \geq 1.5$), "equivocal" for a 30 to 50% increase ($1.5 > F_{max}/F_0 \geq 1.3$), and "nonresponders" if no change >30% was observed ($F_{max}/F_0 < 1.3$). Only cells observed contacting a firing axon were considered in the coculture condition.

For the ratiometric analysis of Di-8-ANNEPS imaging, both channels were background subtracted, and the ratio of channels was generated for each pixel. The resulting image was thresholded and is presented using the "Fire" LUT. A binary mask created from original images that included all neuronal bodies, as well as axons and DCs, was used to assign a zero value to background pixels.

3D reconstruction of confocal images of skin whole mounts and the generation of surfaces were performed in Imaris 9.2.1.

RNA-seq

CD45.1 BMDCs were incubated with nociceptors from CD45.2 mice or alone and treated with the appropriate stimuli. After 8 hours, the cultured cells were harvested, stained, and sorted on a MoFlo Astrios EQ FACS-sorter (Beckman). cDC1s were identified as CD45.1^{+}XCR-1^{+}CD11b^{-}, cDC2s as CD45.1^{+}CD11b^{+}XCR-1^{-}, and pDCs as CD45.1^{+}XCR-1^{-}CD11b^{-}PDCA1^{+}. One thousand cells of each population were sorted directly into 5 μl of lysis buffer (TCL Buffer from Qiagen with 1% β-mercaptoethanol), and Smart-seq2 libraries were prepared as previously described (*74*, *75*) with slight modifications. Briefly, total RNA was captured and purified on RNAClean XP beads (Beckman Coulter). Polyadenylated mRNA was then selected using an anchored oligo(dT) primer (5′-AAGCAGTGGTATCAACGCAGAGTACT30VN-3′) and converted to cDNA through reverse transcription. First-strand cDNA was subjected to limited polymerase chain reaction (PCR) amplification, followed by transposon-based fragmentation using the Nextera XT DNA Library Preparation Kit (Illumina). Samples were PCR amplified for 18 cycles using barcoded primers such that each sample carried a specific combination of eight base Illumina P5 and P7 barcodes, and samples were pooled together before sequencing. Paired-end sequencing was performed on an Illumina NextSeq500 using 2 × 25 bp reads.

RNA-seq data analysis

Reads were aligned to the mouse genome (GENCODE GRCm38/mm10 primary assembly and gene annotations vM16; https://www.gencodegenes.org/mouse/release_M16.html) with STAR 2.5.4a (https://github.com/alexdobin/STAR/releases). The ribosomal RNA gene annotations were removed from the GTF (general transfer format) file. The gene-level quantification was calculated by featureCounts (https://subread.sourceforge.net/). Raw reads count tables were normalized in iDEP.91 {EdgeR [Log2(CPM+1)]} (*76*). Samples with fewer than 2×10^6 uniquely mapped reads were excluded to mitigate the effect of poor-quality samples on normalized counts, and only genes with minimal 50 counts per million (CPM) expression in at least two libraries were selected for further analysis to exclude genes with little to no expression across all conditions. Differentially expressed genes were identified using DESeq2 with FC ≥ 2 and FDR 0.1 cutoffs in iDEP.91.

PCA was performed and visualized using Population PCA software (https://cbdm.hms.harvard.edu/LabMembersPges/SD.html) for 3D PCA, or prcomp function within the R3.6.1 package and visualized using ggbiplot for 2D PCA. Morpheus (https://software.broadinstitute.org/morpheus/) was used for visualization and heatmap generation. GSEA analysis (*37*, *38*) was performed in GSEA 4.0.1 software using the Hallmarks gene matrix v7.1 (h.all.v7.1) (*39*) and the Biocarta gene matrix v7.4 (c2.cp.biocarta.v7.4) MAPK_PATHWAY and P38_MAPK_PATHWAY subsets.

Antibodies

All antibodies used in this study are listed in table S2.

In vivo nociceptor depletion

RTX-mediated depletion of nociceptor innervation was performed as described previously (*9*). Four-week-old mice were injected on 3 consecutive days with 30, 70, and 100 mg/kg body weight of RTX or vehicle control and allowed to age for at least 6 weeks before being used for experiments. Functional denervation was confirmed using the tail withdrawal assay (*9*).

Imiquimod and capsaicin treatment in vivo

Eight- to 20-week-old C57BL/6J mice were treated topically with capsaicin or IMQ as described previously (*9*, *77*). Briefly, for capsaicin treatment, mice were anesthetized, and 250 μg of capsaicin [50 mM in dimethyl sulfoxide (DMSO)] was applied to the treated ear (125 μg on each side) twice 12 hours apart. Pro–IL-1β expression was assessed 24 hours after the first dose. For IMQ treatment, mice were anesthetized, and 25 mg of IMQ in the form of a 5% cream (Aldara) was applied to the treated ear three times 24 hours apart. Where indicated, 3 hours before the IMQ treatment, 5 mg of QX-314 and 5 mg of lidocaine in the form of 4% cream (Aspercreme) was applied to the treated skin. IL-1β, pro–IL-1β, IL-6, or IL-12 p40 expression was assessed 72 hours after the first treatment.

Intradermal injections

Six- to 10-week-old C57BL/6J mice were anesthetized, and 10 μl of PBS, corn oil with 10% DMSO, or corn oil with 10% DMSO and 5 mg/ml olcegepant was injected intradermally into the dorsal side of the ear pinna. After 16 hours, the animals were sacrificed and the injected area of the ear was separated and prepared for flow cytometry analysis.

Skin flow cytometry analysis

Mouse ears were split into the dorsal and ventral aspects, cut into small pieces, and incubated for 40 to 60 min at 37°C with constant shaking in RPMI 1640 medium containing DNase (100 μg/ml) and TM Liberase (62.5 μg/ml) or DNase (100 μg/ml) and collagenase D (2.5 mg/ml). Single-cell suspensions were generated using a GentleMACS dissociator (Miltenyi Biotec), and mechanical disruption was then performed using 50-μm cell strainers. Single-cell suspensions were kept on ice, dead cells were identified using the LIVE/DEAD Fixable Near-IR Dead Cell Stain Kit (Thermo Fisher), Fc receptors were blocked with the anti-CD16/32 antibody, and surface markers were stained with the appropriate fluorescent antibodies in FACS buffer (1% FBS, 5 mM EDTA, and 0.1% NaN_3). Intracellular staining was performed using the Cytofix/Cytoperm kit (BD Biosciences) as per the manufacturer's instructions. DCs were identified as CD45^{+}CD11c^{+}MHC-II^{+}CD64^{-}Ly6C^{-}, monocytes as CD45^{+}CD3^{-}CD19^{-}NK1.1^{-}Ly6C^{hi}Ly6G^{-}, and neutrophils as CD45^{+}CD3^{-}CD19^{-}NK1.1^{-}Ly6C^{lo}Ly6G^{+}.

CHS

CHS experiments were performed as described previously (*78*). Briefly, abdomens of 8- to 12-week-old C57BL/6J mice were shaved and sensitized the following day with 50 μl of 0.5% 2,4-dinitro-1-fluorobenzene (DNFB) in acetone or vehicle control. For experiments with lidocaine + QX314, 5 mg of QX314 and 5 mg of lidocaine in the form of 4% cream (Aspercreme) were applied to the shaved skin 4 hours before sensitization and in 24-hour intervals thereafter for the following 2 days. On day 5, mouse ears were challenged with 20 μl of 0.2% DNFB in acetone or vehicle control, and ear thickness was measured for the following 4 to 5 days. Antigen-specific ear swelling was calculated as DNFB-treated ear thickness minus control ear thickness. The experimenter performing ear-swelling measurements was blinded to mouse genotype and treatment group.

Statistical analysis

Statistical analyses were performed as described in figure legends. Student's t test was used for statistical comparisons of two experimental groups. If more than two experimental groups were present, one-way analysis of variance (ANOVA) with Tukey's multiple-comparisons test was used. For analyses that included two independent variables, two-way ANOVA with Tukey's multiple-comparisons test was used. All statistical analyses were performed in GraphPad Prism version 9.3.1 software.

REFERENCES AND NOTES

1. P. Baral, S. Udit, I. M. Chiu, Pain and immunity: Implications for host defence. *Nat. Rev. Immunol.* **19**, 433–447 (2019). doi: 10.1038/s41577-019-0147-2; pmid: 30874629
2. J. Ordovas-Montanes *et al.*, The regulation of immunological processes by peripheral neurons in homeostasis and disease. *Trends Immunol.* **36**, 578–604 (2015). doi: 10.1016/j.it.2015.08.007; pmid: 26431937
3. S. Huang *et al.*, Lymph nodes are innervated by a unique population of sensory neurons with immunomodulatory potential. *Cell* **184**, 441–459.e25 (2021). doi: 10.1016/j.cell.2020.11.028; pmid: 33333021
4. F. A. Pinho-Ribeiro, W. A. Verri Jr., I. M. Chiu, Nociceptor sensory neuron-immune interactions in pain and inflammation. *Trends Immunol.* **38**, 5–19 (2017). doi: 10.1016/j.it.2016.10.001; pmid: 27793571

5. C. J. Woolf, Q. Ma, Nociceptors—Noxious stimulus detectors. *Neuron* **55**, 353–364 (2007). doi: 10.1016/j.neuron.2007.07.016; pmid: 17678850
6. H. Veiga-Fernandes, D. Mucida, Neuro-immune interactions at barrier surfaces. *Cell* **165**, 801–811 (2016). doi: 10.1016/j.cell.2016.04.041; pmid: 27153494
7. M. Cabeza-Cabrerizo, A. Cardoso, C. M. Minutti, M. Pereira da Costa, C. Reis e Sousa, Dendritic cells revisited. *Annu. Rev. Immunol.* **39**, 131–166 (2021). doi: 10.1146/annurev-immunol-061020-053707; pmid: 33481643
8. S. W. Kashem *et al*., Nociceptive sensory fibers drive interleukin-23 production from CD301b+ dermal dendritic cells and drive protective cutaneous immunity. *Immunity* **43**, 515–526 (2015). doi: 10.1016/j.immuni.2015.08.016; pmid: 26377898
9. L. Riol-Blanco *et al*., Nociceptive sensory neurons drive interleukin-23-mediated psoriasiform skin inflammation. *Nature* **510**, 157–161 (2014). doi: 10.1038/nature13199; pmid: 24759321
10. C. Perner *et al*., Substance P release by sensory neurons triggers dendritic cell migration and initiates the type-2 immune response to allergens. *Immunity* **53**, 1063–1077.e7 (2020). doi: 10.1016/j.immuni.2020.10.001; pmid: 33098765
11. K. Tsujikawa *et al*., Hypertension and dysregulated proinflammatory cytokine production in receptor activity-modifying protein 1-deficient mice. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16702–16707 (2007). doi: 10.1073/pnas.0705974104; pmid: 17923674
12. S. Zelenay, C. Reis e Sousa, Adaptive immunity after cell death. *Trends Immunol.* **34**, 329–335 (2013). doi: 10.1016/j.it.2013.03.005; pmid: 23608152
13. A. M. Binshtok *et al*., Coapplication of lidocaine and the permanently charged sodium channel blocker QX-314 produces a long-lasting nociceptive blockade in rodents. *Anesthesiology* **111**, 127–137 (2009). doi: 10.1097/ALN.0b013e3181a915e7; pmid: 19512868
14. S. J. Gibson *et al*., Plasmacytoid dendritic cells produce cytokines and mature in response to the TLR7 agonists, imiquimod and resiquimod. *Cell. Immunol.* **218**, 74–86 (2002). doi: 10.1016/S0008-8749(02)00517-8; pmid: 12470615
15. H. Hemmi *et al*., Small anti-viral compounds activate immune cells via the TLR7 MyD88-dependent signaling pathway. *Nat. Immunol.* **3**, 196–200 (2002). doi: 10.1038/ni758; pmid: 11812998
16. T. Liu, Z. Z. Xu, C. K. Park, T. Berta, R. R. Ji, Toll-like receptor 7 mediates pruritus. *Nat. Neurosci.* **13**, 1460–1462 (2010). doi: 10.1038/nn.2683; pmid: 21037581
17. J. M. Lund *et al*., Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5598–5603 (2004). doi: 10.1073/pnas.0400937101; pmid: 15034168
18. K. Esancy *et al*., A zebrafish and mouse model for selective pruritus via direct activation of TRPA1. *eLife* **7**, e32036 (2018). doi: 10.7554/eLife.32036; pmid: 29561265
19. J. J. Wei *et al*., Activation of TRPA1 nociceptor promotes systemic adult mammalian skin regeneration. *Sci. Immunol.* **5**, eaba5683 (2020). doi: 10.1126/sciimmunol.aba5683; pmid: 32859683
20. L. Deng, I. M. Chiu, Microbes and pain. *PLOS Pathog.* **17**, e1009398 (2021). doi: 10.1371/journal.ppat.1009398; pmid: 33793668
21. K. Maruyama *et al*., The ATP Transporter VNUT Mediates Induction of Dectin-1-Triggered Candida Nociception. *iScience* **6**, 306–318 (2018). doi: 10.1016/j.isci.2018.08.007; pmid: 30240621
22. V. Meseguer *et al*., TRPA1 channels mediate acute neurogenic inflammation and pain produced by bacterial endotoxins. *Nat. Commun.* **5**, 3125 (2014). doi: 10.1038/ncomms4125; pmid: 24445575
23. B. M. Assas, J. I. Pennock, J. A. Miyan, Calcitonin gene-related peptide is a key neurotransmitter in the neuro-immune axis. *Front. Neurosci.* **8**, 23 (2014). doi: 10.3389/fnins.2014.00023; pmid: 24592205
24. S. Suvas, Role of substance P neuropeptide in inflammation, wound healing, and tissue homeostasis. *J. Immunol.* **199**, 1543–1552 (2017). doi: 10.4049/jimmunol.1601751; pmid: 28827386
25. D. Ganea, M. Delgado, Vasoactive intestinal peptide (VIP) and pituitary adenylate cyclase-activating polypeptide (PACAP) as modulators of both innate and adaptive immunity. *Crit. Rev. Oral Biol. Med.* **13**, 229–237 (2002). doi: 10.1177/154411130201300303; pmid: 12090463
26. B. Lu *et al*., Abnormalities in monocyte recruitment and cytokine expression in monocyte chemoattractant protein 1-deficient mice. *J. Exp. Med.* **187**, 601–608 (1998). doi: 10.1084/jem.187.4.601; pmid: 9463410
27. X. J. Liu *et al*., TLR signaling adaptor protein MyD88 in primary sensory neurons contributes to persistent inflammatory and neuropathic pain and neuroinflammation. *Sci. Rep.* **6**, 28188 (2016). doi: 10.1038/srep28188; pmid: 27312666
28. C. L. Sokol, A. D. Luster, The chemokine system in innate immunity. *Cold Spring Harb. Perspect. Biol.* **7**, a016303 (2015). doi: 10.1101/cshperspect.a016303; pmid: 25635046
29. M. Gschwandtner, R. Derler, K. S. Midwood, More than just attractive: How CCL2 influences myeloid cell behavior beyond chemotaxis. *Front. Immunol.* **10**, 2759 (2019). doi: 10.3389/fimmu.2019.02759; pmid: 31921102
30. A. T. Satpathy *et al*., Zbtb46 expression distinguishes classical dendritic cells and their committed progenitors from other immune lineages. *J. Exp. Med.* **209**, 1135–1152 (2012). doi: 10.1084/jem.20120030; pmid: 22615127
31. M. J. Caterina *et al*., The capsaicin receptor: A heat-activated ion channel in the pain pathway. *Nature* **389**, 816–824 (1997). doi: 10.1038/39807; pmid: 9349813
32. A. M. Taylor *et al*., A microfluidic culture platform for CNS axonal injury, regeneration and transport. *Nat. Methods* **2**, 599–605 (2005). doi: 10.1038/nmeth777; pmid: 16094385
33. Y. L. Huang, A. S. Walker, E. W. Miller, A photostable silicon rhodamine platform for optical voltage sensing. *J. Am. Chem. Soc.* **137**, 10767–10776 (2015). doi: 10.1021/jacs.5b06644; pmid: 26237573
34. E. Shumilina, S. M. Huber, F. Lang, Ca2+ signaling in the regulation of dendritic cell functions. *Am. J. Physiol. Cell Physiol.* **300**, C1205–C1214 (2011). doi: 10.1152/ajpcell.00039.2011; pmid: 21451105
35. N. Matzner *et al*., Ion channels modulating mouse dendritic cell functions. *J. Immunol.* **181**, 6803–6809 (2008). doi: 10.4049/jimmunol.181.10.6803; pmid: 18981098
36. E. Huang, L. Showalter, S. Xu, B. J. Czernliecki, G. K. Koski, Calcium mobilizing treatment acts as a co-signal for TLR-mediated induction of Interleukin-12 (IL-12p70) secretion by murine bone marrow-derived dendritic cells. *Cell. Immunol.* **314**, 26–35 (2017). doi: 10.1016/j.cellimm.2017.01.010; pmid: 28190517
37. V. K. Mootha *et al*., PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003). doi: 10.1038/ng1180; pmid: 12808457
38. A. Subramanian *et al*., Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi: 10.1073/pnas.0506580102; pmid: 16199517
39. A. Liberzon *et al*., The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015). doi: 10.1016/j.cels.2015.12.004; pmid: 26771021
40. S. K. Wculek, S. C. Khouili, E. Priego, I. Heras-Murillo, D. Sancho, Metabolic control of dendritic cell functions: Digesting information. *Front. Immunol.* **10**, 775 (2019). doi: 10.3389/fimmu.2019.00775; pmid: 31073300
41. L. Franchi, T. Eigenbrod, R. Muñoz-Planillo, G. Nuñez, The inflammasome: A caspase-1-activation platform that regulates immune responses and disease pathogenesis. *Nat. Immunol.* **10**, 241–247 (2009). doi: 10.1038/ni.1703; pmid: 19221555
42. T. S. P. Heng *et al*., The Immunological Genome Project: Networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008). doi: 10.1038/ni1008-1091; pmid: 18800157
43. C. S. Walker, A. C. Conner, D. R. Poyner, D. L. Hay, Regulation of signal transduction by calcitonin gene-related peptide receptors. *Trends Pharmacol. Sci.* **31**, 476–483 (2010). doi: 10.1016/j.tips.2010.06.006; pmid: 20633935
44. Z. G. Goldsmith, D. N. Dhanasekaran, G protein regulation of MAPK networks. *Oncogene* **26**, 3122–3142 (2007). doi: 10.1038/sj.onc.1210407; pmid: 17496911
45. C. Perner, C. L. Sokol, Protocol for dissection and culture of murine dorsal root ganglia neurons to study neuropeptide release. *STAR Protoc.* **2**, 100333 (2021). doi: 10.1016/j.xpro.2021.100333; pmid: 33615276
46. F. A. Pinho-Ribeiro *et al*., Blocking neuronal signaling to immune cells treats streptococcal invasive infection. *Cell* **173**, 1083–1097.e22 (2018). doi: 10.1016/j.cell.2018.04.006; pmid: 29754819
47. S. Ständer *et al*., Expression of vanilloid receptor subtype 1 in cutaneous sensory nerve fibers, mast cells, and epithelial cells of appendage structures. *Exp. Dermatol.* **13**, 129–139 (2004). doi: 10.1111/j.0906-6705.2004.0178.x; pmid: 14987252
48. Z. Zasłona *et al*., The Induction of Pro-IL-1β by Lipopolysaccharide Requires Endogenous Prostaglandin E_2 Production. *J. Immunol.* **198**, 3558–3564 (2017). doi: 10.4049/jimmunol.1602072; pmid: 28298525
49. C. Shi *et al*., Bone marrow mesenchymal stem and progenitor cells induce monocyte emigration in response to circulating toll-like receptor ligands. *Immunity* **34**, 590–601 (2011). doi: 10.1016/j.immuni.2011.02.016; pmid: 21458307
50. A. D. Christensen, C. Haase, Immunological mechanisms of contact hypersensitivity in mice. *APMIS* **120**, 1–27 (2012). doi: 10.1111/j.1600-0463.2011.02832.x; pmid: 22151305
51. J. Filtjens *et al*., Nociceptive sensory neurons promote CD8 T cell responses to HSV-1 infection. *Nat. Commun.* **12**, 2936 (2021). doi: 10.1038/s41467-021-22841-6; pmid: 34006861
52. J. Liu, M. Chen, X. Wang, Calcitonin gene-related peptide inhibits lipopolysaccharide-induced interleukin-12 release from mouse peritoneal macrophages, mediated by the cAMP pathway. *Immunology* **101**, 61–67 (2000). doi: 10.1046/j.1365-2567.2000.00082.x; pmid: 11012754
53. I. M. Chiu *et al*., Bacteria activate sensory neurons that modulate pain and inflammation. *Nature* **501**, 52–57 (2013). doi: 10.1038/nature12479; pmid: 23965627
54. P. Baral *et al*., Nociceptor sensory neurons suppress neutrophil and γδ T cell responses in bacterial lung infections and lethal pneumonia. *Nat. Med.* **24**, 417–426 (2018). doi: 10.1038/nm.4501; pmid: 29505031
55. H. Nagashima *et al*., Neuropeptide CGRP limits group 2 innate lymphoid cell responses and constrains type 2 inflammation. *Immunity* **51**, 682–695.e6 (2019). doi: 10.1016/j.immuni.2019.06.009; pmid: 31353223
56. J. Helft *et al*., GM-CSF mouse bone marrow cultures comprise a heterogeneous population of CD11c(+)MHCII(+) macrophages and dendritic cells. *Immunity* **42**, 1197–1211 (2015). doi: 10.1016/j.immuni.2015.05.018; pmid: 26084029
57. M. Baliu-Piqué, G. Jusek, B. Holzmann, Neuroimmunological communication via CGRP promotes the development of a regulatory phenotype in TLR4-stimulated macrophages. *Eur. J. Immunol.* **44**, 3708–3716 (2014). doi: 10.1002/eji.201444553; pmid: 25316186
58. F. L. Liu *et al*., Interleukin (IL)-23 p19 expression induced by IL-1beta in human fibroblast-like synoviocytes with rheumatoid arthritis via active nuclear factor-kappaB and AP-1 dependent pathway. *Rheumatology (Oxford)* **46**, 1266–1273 (2007). doi: 10.1093/rheumatology/kem055; pmid: 17569750
59. A. Malecka *et al*., Stromal fibroblasts support dendritic cells to maintain IL-23/Th17 responses after exposure to ionizing radiation. *J. Leukoc. Biol.* **100**, 381–389 (2016). doi: 10.1189/jlb.3A1015-474R; pmid: 27049023
60. O. Chen, C. R. Donnelly, R. R. Ji, Regulation of pain by neuro-immune interactions between macrophages and nociceptor sensory neurons. *Curr. Opin. Neurobiol.* **62**, 17–25 (2020). doi: 10.1016/j.conb.2019.11.006; pmid: 31809997
61. J. Kupari *et al*., Single cell transcriptomics of primate sensory neurons identifies cell types associated with chronic pain. *Nat. Commun.* **12**, 1510 (2021). doi: 10.1038/s41467-021-21725-z; pmid: 33686078
62. D. Usoskin *et al*., Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* **18**, 145–153 (2015). doi: 10.1038/nn.3881; pmid: 25420068
63. X. Wang, M. L. Veruki, N. V. Bukoreshtliev, E. Hartveit, H. H. Gerdes, Animal cells connected by nanotubes can be electrically coupled through interposed gap-junction channels. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17194–17199 (2010). doi: 10.1073/pnas.1006785107; pmid: 20855598
64. J. Ariazi *et al*., Tunneling nanotubes and gap junctions-their role in long-range intercellular communication during development, health, and disease conditions. *Front. Mol. Neurosci.* **10**, 333 (2017). doi: 10.3389/fnmol.2017.00333; pmid: 29089870
65. M. Krummen *et al*., Release of IL-12 by dendritic cells activated by TLR ligation is dependent on MyD88 signaling, whereas TRIF signaling is indispensable for TLR synergy. *J. Leukoc. Biol.* **88**, 189–199 (2010). doi: 10.1189/jlb.0408228; pmid: 20360404
66. M. Boes *et al*., T-cell engagement of dendritic cells rapidly rearranges MHC class II transport. *Nature* **418**, 983–988 (2002). doi: 10.1038/nature01004; pmid: 12198548
67. M. A. Nassar *et al*., Nociceptor-specific gene deletion reveals a major role for Nav1.7 (PN1) in acute and inflammatory pain. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12706–12711 (2004). doi: 10.1073/pnas.0404915101; pmid: 15314237
68. K. Brasel, T. De Smedt, J. L. Smith, C. R. Maliszewski, Generation of murine dendritic cells from flt3-ligand-supplemented bone marrow cultures. *Blood* **96**, 3029–3039 (2000). doi: 10.1182/blood.V96.9.3029; pmid: 11049981

69. J. Weischenfeldt, B. Porse, Bone marrow-derived macrophages (BMM): Isolation and applications. *CSH Protoc.* **2008**, pdb.prot5080 (2008). doi: 10.1101/pdb.prot5080; pmid: 21356739
70. S. Katzenell, J. R. Cabrera, B. J. North, D. A. Leib, Isolation, purification, and culture of primary murine sensory neurons. *Methods Mol. Biol.* **1656**, 229–251 (2017). doi: 10.1007/978-1-4939-7237-1_15; pmid: 28808974
71. T. Fath, Y. D. Ke, P. Gunning, J. Götz, L. M. Ittner, Primary support cultures of hippocampal and substantia nigra neurons. *Nat. Protoc.* **4**, 78–85 (2009). doi: 10.1038/nprot.2008.199; pmid: 19131959
72. J. Schindelin *et al*., Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012). doi: 10.1038/nmeth.2019; pmid: 22743772
73. B. C. Chen *et al*., Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 1257998 (2014). doi: 10.1126/science.1257998; pmid: 25342811
74. S. Picelli *et al*., Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013). doi: 10.1038/nmeth.2639; pmid: 24056875
75. S. Picelli *et al*., Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014). doi: 10.1038/nprot.2014.006; pmid: 24385147
76. S. X. Ge, E. W. Son, R. Yao, iDEP: An integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* **19**, 534 (2018). doi: 10.1186/s12859-018-2486-6; pmid: 30567491
77. H. Inoue, N. Nagata, Y. Koshihara, Participation of serotonin in capsaicin-induced mouse ear edema. *Jpn. J. Pharmacol.* **69**, 61–68 (1995). doi: 10.1254/jjp.69.61; pmid: 8847833
78. S. Paust *et al*., Critical role for the chemokine receptor CXCR6 in NK cell-mediated antigen-specific memory of haptens and viruses. *Nat. Immunol.* **11**, 1127–1135 (2010). doi: 10.1038/ni.1953; pmid: 20972432

ACKNOWLEDGMENTS

We thank H. Ploegh for sharing MHC-II eGFP mice; members of the von Andrian laboratory for helpful advice and discussions; and Harvard Medical School Electron Microscopy, Flow Cytometry, Microscopy Resources on the Northern Quad (MicRoN) and Cell Biology Microscopy facilities for technical assistance. Influenza PR8-GFP was generously provided by A. Garcia-Sastre at the Mount Sinai School of Medicine and the NIAID Centers of Excellence for Influenza Research and Surveillance program. *Streptococcus pneumoniae* was a kind gift from R. Malley at the Boston Children's Hospital and Harvard Medical School. **Funding:** This work was supported by the National Institutes of Health (grant AR068383 to U.H.v.A. and grant R01NS098088 to E.W.M.), a CRI Irvington postdoctoral fellowship (CRI2453 to P.H.), and the Howard Hughes Medical Institute (Gilliam Fellowship to G.O.). **Author contributions:** P.H. and U.H.v.A. conceived the study, interpreted the results, and wrote the manuscript. P.H. performed experiments and analyzed data. R.J.G. performed intradermal injections. I.B.M. generated BM chimeras and performed CHS experiments. Y.W. prepared cortical and hippocampal neuronal cultures. T.L. performed light sheet microscopy imaging. G.O. and E.W.M. generated the voltage-sensitive dye BeRST. **Competing interests:** U.H.v.A. is a paid consultant with financial interests in Avenge Bio, Beam Therapeutics, Bluesphere Bio, Curon, DNAlite, Gate Biosciences, Gentibio, Intergalactic, intrECate Biotherapeutics, Interon, Mallinckrodt Pharmaceuticals, Moderna, Monopteros Biotherapeutics, Morphic Therapeutics, Rubius, Selecta and SQZ. E.W.M. is an inventor on a patent describing BeRST-WO2017019908. The remaining authors declare no competing interests. **Data and materials availability:** RNA-seq data are available in the Gene Expression Omnibus (GEO) under GSE 217503, subseries GSE 217500 (data shown in Fig. 5) and GSE 217502 (data shown in Fig. 6C). All remaining data are available in the main text or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm5658
Figs. S1 to S14
Tables S1 and S2
Movies S1 to S10
References (*79–113*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 26 September 2021; resubmitted 12 August 2022
Accepted 17 February 2023
10.1126/science.abm5658

Check for updates

RESEARCH ARTICLE SUMMARY

PLANT SCIENCE

Brassinosteroid gene regulatory networks at cellular resolution in the *Arabidopsis* root

Trevor M. Nolan†, Nemanja Vukašinović†, Che-Wei Hsu†, Jingyuan Zhang, Isabelle Vanhoutte, Rachel Shahan, Isaiah W. Taylor, Laura Greenstreet, Matthieu Heitz, Anton Afanassiev, Ping Wang, Pablo Szekely, Aiden Brosnan, Yanhai Yin, Geoffrey Schiebinger, Uwe Ohler, Eugenia Russinova*, Philip N. Benfey*

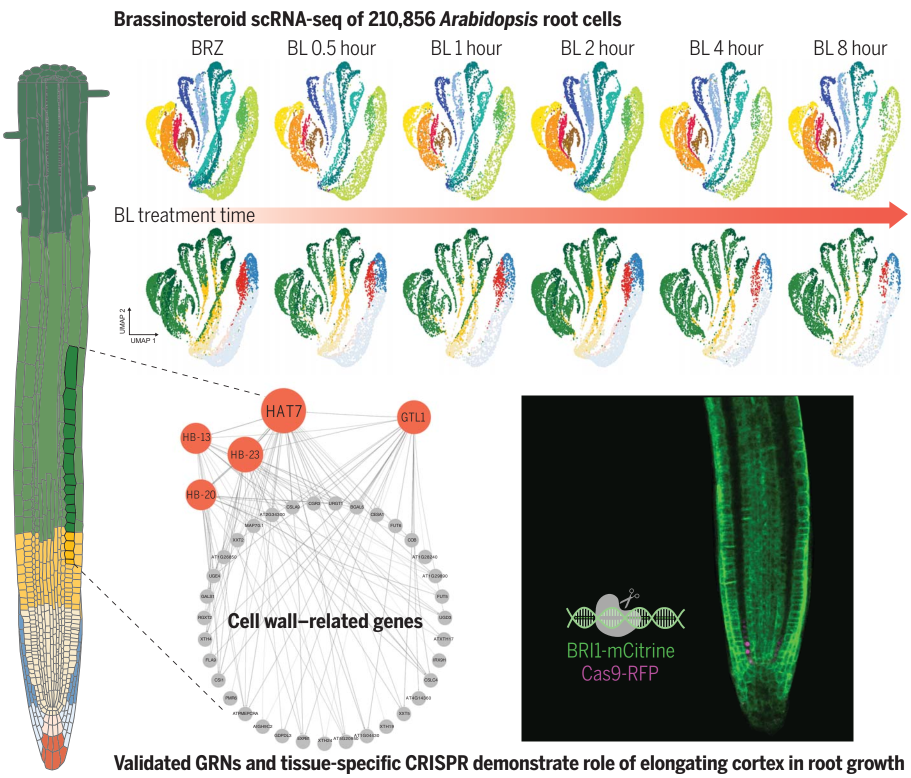
INTRODUCTION: Cells traverse a developmental landscape as they acquire identities and progress toward end-stage differentiation. Gene regulatory networks control this progression and must be tuned according to developmental stage, cell identity, and environmental conditions. Hormones play important roles in remodeling these networks, but it has been challenging to understand how cell identities, developmental states, and hormone responses influence one another. Brassinosteroids are plant steroid hormones that regulate diverse processes, including cell division and cell elongation. Brassinosteroids signal to activate BRI1-EMS-SUPPRESSOR 1 (BES1) and BRASSINAZOLE-RESISTANT 1 (BZR1) transcription factors, which direct gene regulatory networks to control thousands of genes. Modulating brassinosteroids can lead to different responses depending on the developmental context, but how the underlying gene regulatory networks vary in space and time is unclear.

RATIONALE: Single-cell RNA sequencing (scRNA-seq) is a powerful approach to investigate cell- and developmental stage–specific responses to stimuli, but most previous studies have focused on a single time point. In this work, we used time series scRNA-seq to delineate the gene regulatory networks controlling brassinosteroid response in the *Arabidopsis* root. We then confirmed the spatial and developmental models arising from single-cell analysis using tissue-specific gene manipulations.

RESULTS: We defined brassinosteroid-responsive gene expression using time series scRNA-seq. This identified the elongating root cortex as a site of brassinosteroid-mediated gene expression. Reconstruction of cortex trajectories showed that brassinosteroids promote a shift from proliferation to elongation associated with increased expression of cell wall–related genes. Accordingly, loss of brassinosteroid signaling in the cortex using a tissue-specific CRISPR approach impaired cell expansion in the elongation zone but had little effect on meristem cell length.

To discover regulators of spatiotemporal brassinosteroid responses, we inferred gene regulatory networks across each cell type, developmental stage, and time point of our brassinosteroid time series. Our gene regulatory networks and experimental analysis revealed *HOMEOBOX FROM ARABIDOPSIS THALIANA 7* (*HAT7*) and *GT-2-LIKE 1* (*GTL1*) as brassinosteroid-responsive transcription factors that regulate cell elongation in the cortex. BES1 and GTL1 interact and control a common set of targets induced by brassinosteroids, as evidenced by misregulation in scRNA-seq of *gtl1 df1* mutants. These datasets represent 210,856 single-cell transcriptomes, providing a high-resolution view of brassinosteroid-mediated gene regulatory networks.

CONCLUSION: We have established the cortex as a site for brassinosteroid-mediated gene expression, where brassinosteroids activate cell wall–related genes and promote elongation. We further showed that *HAT7* and *GTL1* are brassinosteroid-induced regulators along cortex trajectories that control cell elongation. These findings highlight the ability of scRNA-seq to identify context-specific transcription factors, which could be leveraged to precisely engineer plant growth and development. Our results unveil a brassinosteroid signaling network regulating the transition from proliferation to elongation in the cortex, illuminating a spatiotemporal brassinosteroid response. ■



Brassinosteroid networks controlling root cell elongation. scRNA-seq of 210,856 *Arabidopsis* root cells from brassinosteroid treatments and mutants identified the elongating cortex as a site of brassinosteroid response. Gene regulatory networks involving HAT7 and GTL1 family transcription factors induce cell wall–related genes as cells transition to elongation. Tissue-specific CRISPR of *BRASSINOSTEROID INSENSITIVE 1* (*BRI1*) demonstrated the role of the cortex in brassinosteroid-mediated cell elongation. BRZ, Brassinazole; BL, Brassinolide; UMAP, uniform manifold approximation and projection; RFP, red fluorescent protein.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: eugenia.russinova@psb.vib-ugent.be (E.R.); philip.benfey@duke.edu (P.N.B.)
†These authors contributed equally to this work.
Cite this article as T. M. Nolan *et al.*, *Science* 379, eadf4721 (2023). DOI: 10.1126/science.adf4721

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adf4721

RESEARCH ARTICLE

PLANT SCIENCE

Brassinosteroid gene regulatory networks at cellular resolution in the *Arabidopsis* root

Trevor M. Nolan[1]†, Nemanja Vukašinović[2,3]†, Che-Wei Hsu[1,4,5]†, Jingyuan Zhang[1], Isabelle Vanhoutte[2,3], Rachel Shahan[1,6], Isaiah W. Taylor[1], Laura Greenstreet[7], Matthieu Heitz[7], Anton Afanassiev[7], Ping Wang[8], Pablo Szekely[1,6], Aiden Brosnan[1], Yanhai Yin[8], Geoffrey Schiebinger[7], Uwe Ohler[4,8,9], Eugenia Russinova[2,3]*, Philip N. Benfey[1,6]*

Brassinosteroids are plant steroid hormones that regulate diverse processes, such as cell division and cell elongation, through gene regulatory networks that vary in space and time. By using time series single-cell RNA sequencing to profile brassinosteroid-responsive gene expression specific to different cell types and developmental stages of the *Arabidopsis* root, we identified the elongating cortex as a site where brassinosteroids trigger a shift from proliferation to elongation associated with increased expression of cell wall–related genes. Our analysis revealed *HOMEOBOX FROM ARABIDOPSIS THALIANA 7* (*HAT7*) and *GT-2-LIKE 1* (*GTL1*) as brassinosteroid-responsive transcription factors that regulate cortex cell elongation. These results establish the cortex as a site of brassinosteroid-mediated growth and unveil a brassinosteroid signaling network regulating the transition from proliferation to elongation, which illuminates aspects of spatiotemporal hormone responses.

During development, cells pass through different states as they acquire identities and progress toward end-stage differentiation (1). Gene regulatory networks (GRNs) control this progression and must be tuned according to developmental stage, cell identity, and environmental conditions (2–4). Signaling molecules, such as hormones, are central players in coordinating these networks, but it has been challenging to disentangle how cell identities, developmental states, and hormone responses influence one another. Technological advances in single-cell RNA sequencing (scRNA-seq) (2, 5) and tissue-specific gene manipulations (6, 7) make it possible to address this challenge using the *Arabidopsis* root as a model system.

Brassinosteroids are a group of plant steroid hormones that affect cell division and cell elongation during root growth (8–12). Brassinosteroids are sensed at the plasma membrane by BRASSINOSTEROID INSENSITIVE 1 (BRI1) family receptors (13–16), initiating signal transduction events that activate BES1 and BZR1 family transcription factors to control thousands of genes (17–20). The brassinosteroid GRN is typically represented singularly without consideration of cell specificity (20–23), even though brassinosteroids lead to different responses depending on the developmental context (24–28).

By profiling brassinosteroid responses across the cell types and developmental stages of the root using scRNA-seq, we discovered that brassinosteroids affect gene expression in the elongating cortex. Reconstruction of cortex trajectories over a scRNA-seq time course showed that brassinosteroids trigger a shift from proliferation to elongation, which is associated with up-regulation of cell wall–related genes. Loss of brassinosteroid signaling in the cortex using tissue-specific CRISPR reduced cell elongation. Our time course data allowed us to propose brassinosteroid-responsive GRNs, which led to the identification of HAT7 and GTL1 as validated regulators of brassinosteroid response in the elongating cortex. These datasets represent 210,856 single-cell transcriptomes, providing a high-resolution view of brassinosteroid-mediated GRNs.

[1]Department of Biology, Duke University, Durham, NC, USA. [2]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium. [3]Center for Plant Systems Biology, VIB, Ghent, Belgium. [4]Department of Biology, Humboldt Universitat zu Berlin, Berlin, Germany. [5]The Berlin Institute for Medical Systems Biology, Max Delbruck Center for Molecular Medicine, Berlin, Germany. [6]Howard Hughes Medical Institute, Duke University, Durham, NC, USA. [7]Department of Mathematics, University of British Columbia, Vancouver, BC, Canada. [8]Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA, USA. [9]Department of Computer Science, Humboldt Universitat zu Berlin, Berlin, Germany.
***Corresponding author. Email: eugenia.russinova@psb.vib-ugent.be (E.R.); philip.benfey@duke.edu (P.N.B.)**
†These authors contributed equally to this work.

scRNA-seq reveals spatiotemporal brassinosteroid responses

To investigate spatiotemporal brassinosteroid responses in the *Arabidopsis* root, we used a sensitized system, which involved inhibiting brassinosteroid biosynthesis using Brassinazole (BRZ) (29) and then reactivating signaling with Brassinolide (BL), the most active brassinosteroid (10, 21, 30) (fig. S1). We treated 7-day-old primary roots for 2 hours with BL or a corresponding mock BRZ control and performed scRNA-seq on protoplasts isolated from three biological replicates of 0.5-cm root tips (containing meristem, elongation, and early differentiation zones) using the 10X Genomics Chromium system (see materials and methods).

To annotate cell types and developmental stages, we performed label transfer based on our single-cell atlas of the *Arabidopsis* root (31). We distinguished between two domains of the meristem: the proliferation domain, where cells have a high probability of dividing, and the transition domain, where cells divide less frequently but have not yet begun rapid expansion (fig. S2, A to D, and data S2) (32, 33).

After data integration, the 11 major cell types and eight developmental stages identified were logically arranged in two-dimensional (2D) uniform manifold approximation and projection (UMAP) space as previously described for root datasets (fig. S3, A and B) (31, 34). Marker genes characteristic of cell types and developmental stages remained enriched, which suggests that although brassinosteroids alter the expression of thousands of genes, cell identities can be successfully aligned through integration (fig. S3C).

Previous studies have profiled brassinosteroid-responsive gene expression in bulk tissue or in a handful of cell types, conflating cell type and developmental stage (23, 27, 35). To obtain a better spatiotemporal resolution, we performed differential expression analysis for each combination of cell type and developmental stage using pseudobulk expression profiles (materials and methods). We identified 8286 differentially expressed genes (DEGs) (fold-change > 1.5, false discovery rate < 0.05; Fig. 1A and fig. S4, A and B), which were enriched in BES1 and BZR1 targets and had significant overlap with previously identified brassinosteroid-regulated genes (fig. S4C and data S3).

We found that 37% of DEGs were altered in a single cell type or developmental stage and >82% were differentially expressed in five or fewer cell type–developmental stage combinations (fig. S4B). This indicates that although brassinosteroids broadly influence gene expression, they modulate distinct sets of genes in different spatiotemporal contexts.

Among the tissues with many DEGs was the epidermis, as previously described (9, 12, 24, 27, 35). Atrichoblasts, or nonhair cells in the epidermis, were particularly affected, showing changes across both the meristem and elongation zone. Our data also indicated that brassinosteroids influence gene expression in the cortex, especially in the elongation zone (Fig. 1A and fig. S4A). The cortex has been linked to plant environmental interactions, including response to water limitation (36, 37) and hydrotropism (38, 39). However, it is unknown how brassinosteroids modulate gene expression in this tissue or what processes are affected. To address

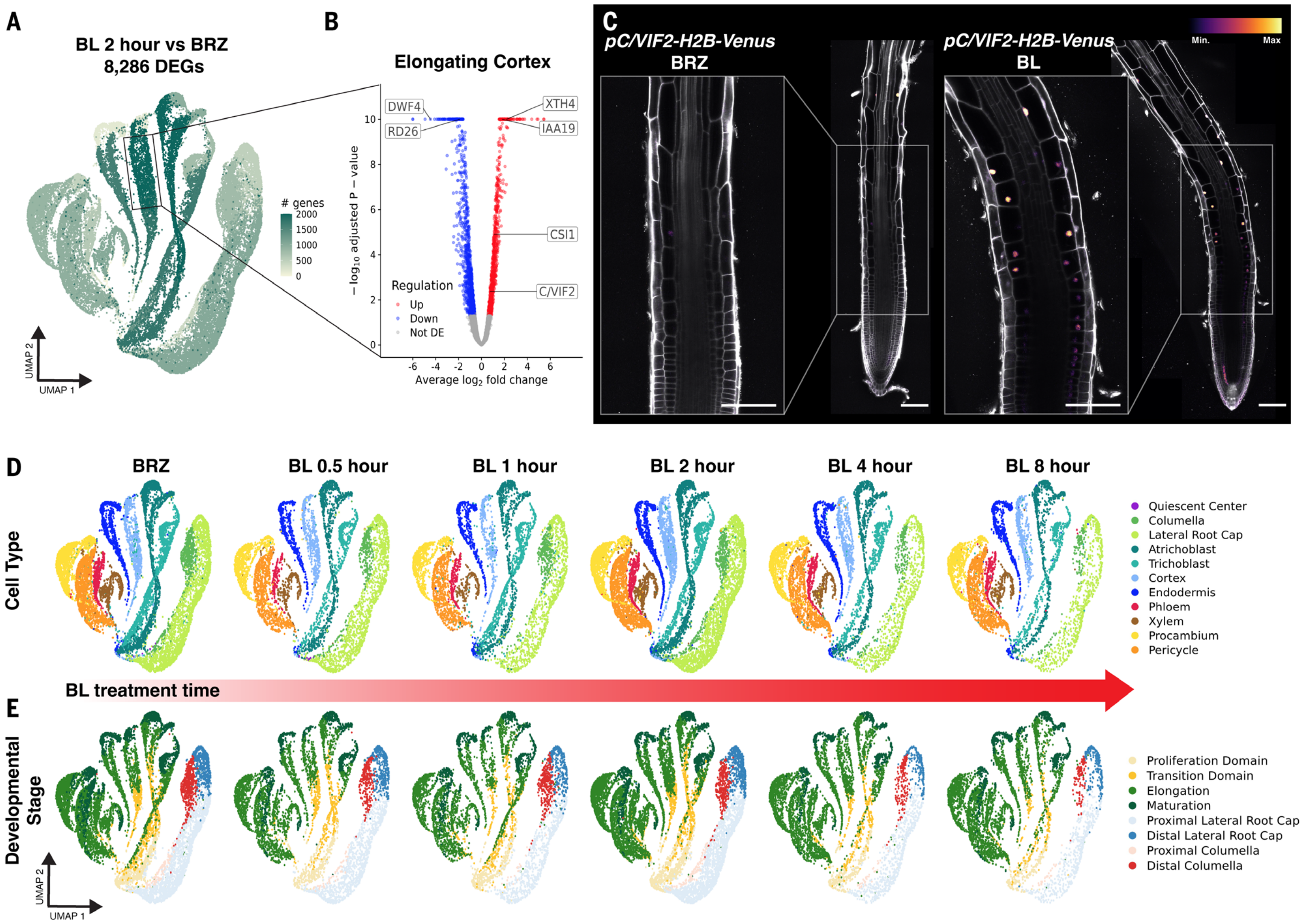


Fig. 1. scRNA-seq identifies brassinosteroid induction of cell wall–related genes in the cortex. (**A**) Spatiotemporal response to 2-hour BL treatment versus BRZ control among each combination of cell type and developmental stage of the *Arabidopsis* root. Color on UMAP projection indicates the number of DEGs. (**B**) Volcano plot of BL DEGs in the elongating cortex. Color indicates the direction of regulation. Known markers of brassinosteroid response, including *DWF4*, *RD26*, *XTH4*, and *IAA19*, are indicated. *C/VIF2* and *CSI1* (described in this study) are also indicated. (**C**) *pC/VIF2-H2B-Venus* reporter grown on 1 μM BRZ for 7 days and transferred to 1 μM BRZ or 100 nM BL for 4 hours. (Inset) *C/VIF2* signals in the elongating cortex that increase with BL treatment. Propidium iodide staining is shown in gray, with the color gradient indicating relative *C/VIF2-H2B-Venus* levels. Scale bars, 100 μm. (**D** and **E**) UMAP of BL treatment scRNA-seq time course. Mock BRZ control represents time 0. Colors indicate cell type (D) or developmental stage annotation (E).

these questions, we focused on brassinosteroid-mediated gene expression in the elongating cortex.

Brassinosteroids induce cell wall–related genes in the elongating cortex

We found that brassinosteroid treatment led to 967 up-regulated genes and 1156 down-regulated genes in the elongating cortex (Fig. 1B and fig. S4A). Gene ontology (GO) analysis indicated that the up-regulated genes were enriched for genes related to cell wall organization or biogenesis, which is intriguing given the role of brassinosteroids in promoting cell elongation (fig. S4D). The cell wall–related DEGs included *CELLULOSE SYNTHASES* (*CESAs*), *CELLULOSE SYNTHASE INTERACTIVE1* (*CSI1*), and cell wall–loosening enzymes such as *EXPANSINS* and *XYLOGLUCAN ENDOTRANSGLUCOSYLASES*. Cell wall–related genes such as *CESAs* have been demonstrated to be direct targets of BES1 and BZR1 (*19*, *20*, *40*), but their spatiotemporal regulation, especially in the cortex, has not been reported.

To monitor their responsiveness to brassinosteroids, we generated transcriptional reporters for three of the DEGs with distinct spatiotemporal patterns (Fig. 1C and fig. S5, A to E). For example, *CELL WALL/VACUOLAR INHIBITOR OF FRUCTOSIDASE 2* (*C/VIF2*) was enriched in the transition domain and elongation zone of the cortex and induced by BL (Fig. 1C). These results confirm that our differential expression analysis captures spatiotemporal brassinosteroid responses and raise the possibility that brassinosteroid induction of cell wall–related genes is associated with cortex cell elongation.

Induction of cell wall–related genes is associated with switch to elongation

To better understand how brassinosteroids influence cell wall–related gene expression, we performed scRNA-seq at six time points beginning with BRZ treatment (time 0) and BL treatments for 30 min, 1 hour, 2 hours, 4 hours, and 8 hours (Fig. 1, D and E). These time points capture the rapid root elongation triggered by the readdition of brassinosteroids (*35*).

Waddington–optimal transport (WOT) is an analytical approach for analyzing expression trends over a scRNA-seq time course. WOT connects snapshots of gene expression to facilitate trajectory reconstruction, identifying putative ancestors for a given set of cells at earlier time points and descendants at later time points (*41*).

To examine the trajectories leading to the activation of cell wall–related genes in the elongating cortex, we applied WOT (*41*) and created a cell wall gene signature using 107 cell wall–related genes that were induced by BL in the elongating cortex. We monitored the relative expression of these genes, resulting in a cell wall score for each cell in the time course (materials and methods). Cortex cells had a higher cell wall score compared with other cell types, which increased with BL treatment (Fig. 2, A and B, and fig. S6, A and B), confirming that the cell wall score represents a brassinosteroid-responsive module in the cortex. At the 2-hour BL time point, >20% of cortex cells had a cell wall score greater than 1. By contrast, only 5% or fewer cells in other cell types exhibited scores this high (Fig. 2B). We

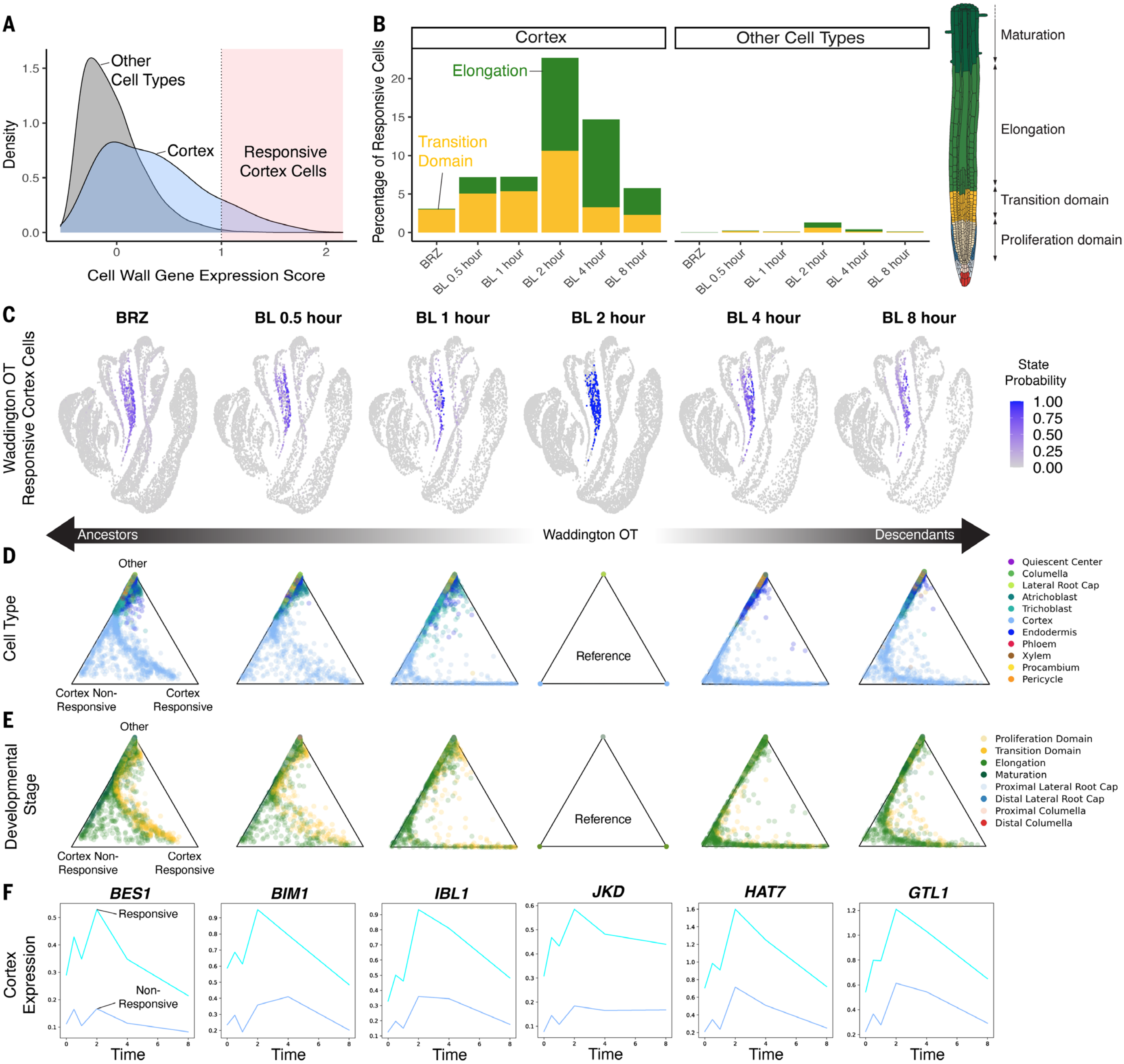


Fig. 2. WOT traces the induction of cell wall–related genes along cortex trajectories associated with the switch to elongation. (**A**) Density plot showing cell wall gene expression score. The shaded region with cell wall expression scores ≥1 indicates responsive cortex cells. (**B**) Bar plot showing the percentage of responsive cells in the cortex versus other cell types over the time course. Color indicates developmental stage annotation, also depicted in the root schematic. Illustration adapted from the Plant Illustrations repository. Only transition and elongation zones are plotted because other zones represent <2% of responsive cells. (**C**) WOT probabilities for cortex responsive state along the BL time course. The BL 2-hour time point was used as a reference; therefore, all cells have a probability of either 1 or 0 at this time point. (**D** and **E**) Triangle plots with cells plotted according to WOT cortex responsive, cortex nonresponsive, or other state probabilities for each time point along the BL time course. Color indicates the cell type (D) or developmental stage (E) annotation. (**F**) Expression trends for select transcription factors differentially expressed along WOT cortex responsive trajectories.

therefore designated cells with a cell wall score of at least 1 as responsive cortex cells to indicate their exceptional brassinosteroid response (Fig. 2A).

An advantage of WOT analysis is that it does not rely on prespecified boundaries between developmental zones. We used this property to examine the relationship between developmental stage annotation and cell wall score. Under BRZ treatment, responsive cortex cells were sparse and predominantly annotated as transition domain. Upon BL treatment, the annotation of responsive cortex cells shifted to the elongation zone (Fig. 2B). Using the cells at the 2-hour time point as a reference, we looked at the probability of cells being ancestors or descendants of responsive cortex cells (Fig. 2C). We also constructed a similar trajectory for the remaining cortex cells, which were designated nonresponsive cortex cells. We visualized these probabilities on barycentric coordinates as triangle plots, where vertices represent a 100% chance of becoming the labeled state and interior positions indicate intermediate probabilities. Cells can be colored according to gene expression, cell types, or other quantities. The cortex nonresponsive state was represented by the lower left vertex, the cortex responsive state was represented by the lower right vertex, and the top vertex represented other states. This illustrated that cortex transition domain cells were predisposed toward the responsive state under BRZ conditions and subsequently shifted from transition domain to elongation zone annotation upon BL treatment (Fig. 2, D and E). These results suggest that brassinosteroids are involved in initiating the elongation of cortex cells through the activation of cell wall–related genes.

WOT trajectories identify brassinosteroid-responsive transcription factors

To reveal potential regulators of cell wall–related genes in the cortex, we performed probabilistic differential expression analysis along WOT trajectories, contrasting cells assigned to cortex responsive versus nonresponsive states at each time point (materials and methods). Among the DEGs identified were known transcription factors in the brassinosteroid pathway, including *BES1* (*17*), *BES1-INTERACTING MYC-LIKE1* (*BIM1*) (*42*), and *IBH1-LIKE 1* (*IBL1*) (*43*) (Fig. 2F). We also identified additional transcription factors, including *JACKDAW* (*JKD*), which is involved in ground tissue specification (*44*); the class I HD-ZIP transcription factor *HAT7* (*45*, *46*); and *GTL1*. Because *JKD* was identified by WOT and a previous BL RNA-seq dataset (*21*) but not pseudobulk differential expression analysis, we confirmed its brassinosteroid responsiveness (fig. S6, C and D). These results indicate that WOT trajectories can identify brassinosteroid-responsive transcription factors that may be involved in regulating cell wall–related genes in the cortex.

Analysis of the triple receptor mutant *bri1-T* reveals changes in cortex expression

Because our results indicated that exogenous brassinosteroids lead to the activation of cell wall–related genes in the elongating cortex, we investigated whether this is also the case for endogenous brassinosteroids. A gradient of brassinosteroids is present along the longitudinal axis of the root, with low brassinosteroid levels in the proliferation domain (*47*). Brassinosteroid biosynthesis increases as cells enter the transition domain and peaks in the elongation zone, shootward of which is a brassinosteroid signaling maximum (*35*, *47*). Interpretation of this endogenous brassinosteroid gradient requires receptor BRI1 and its homologs BRL1 and BRL3 (*13*, *14*, *48*–*50*).

To identify differentially expressed genes, we performed two replicates of scRNA-seq on the brassinosteroid-blind *bri1brl1brl3* triple mutant (*bri1-T*) along with paired wild-type (WT) controls (Fig. 3, A to C). A previous study had profiled single cells from *bri1-T*, suggesting potential brassinosteroid responsiveness of the cortex (*51*). However, these data were from a single replicate, were compared with a WT sample from a different study, and did not resolve developmental stage specificity. By contrast, our analysis across both cell types and developmental stages identified the elongating cortex as exhibiting differential gene expression (Fig. 3, D and E; fig. S7A; and data S3). The genes down-regulated in the elongating cortex of *bri1-T* were enriched for the GO term cell wall organization or biogenesis (fig. S7B). BL treatment increased the proportion of elongating cortex cells (fig. S7C), but fewer elongating cortex cells were observed in *bri1-T* compared with WT (fig. S7D).

To further verify these changes in gene expression, we performed an additional side-by-side scRNA-seq experiment in which we inhibited endogenous brassinosteroids through BRZ treatment. We profiled 16,642 WT control and 14,320 WT BRZ-treated cells and detected 6928 DEGs in BRZ versus control using pseudobulk differential expression analysis of each cell type–developmental stage combination (fig. S8, A to C). Consistent with the idea that brassinosteroids promote cell wall–related gene expression in the elongating cortex, BRZ down-regulated genes in the elongating cortex were enriched for genes associated with cell wall organization or biogenesis (fig. S8D). Together, our analyses of *bri1-T* and BRZ treatment indicate that endogenous brassinosteroid signaling promotes the expression of cell wall–related genes in the cortex associated with the onset of elongation.

The epidermis is widely described as the major site for brassinosteroid-promoted gene expression in the root (*9*, *10*, *24*, *27*, *35*, *52*). Previous studies have shown that epidermal expression of BRI1 was sufficient to rescue morphological phenotypes, including meristem size of *bri1-T* (*9*, *24*, *53*). To determine the extent to which brassinosteroid-regulated gene expression is restored, we performed scRNA-seq on *pGL2-BRI1-GFP/bri1-T* (GFP, green fluorescent protein)—a line in which BRI1 is expressed in atrichoblast cells of the epidermis of *bri1-T* (*9*, *27*, *53*). We identified 8188 DEGs in comparison with WT (Fig. 3F) and 8046 DEGs in comparison with *bri1-T* (Fig. 3G and fig. S9, A to D), which indicates that gene expression remains perturbed (fig. S10, A to D) and that this is far from a complete rescue of the *bri1-T* phenotype.

Tissue-specific CRISPR confirms role of cortex in brassinosteroid-mediated cell expansion

Characterization of cell type–specific brassinosteroid signaling has relied on tissue-specific complementation lines, which has led to conflicting results and has overlooked the role of brassinosteroid signaling in the cortex (*9*, *24*, *27*, *35*, *51*, *53*, *54*). To selectively block brassinosteroid signaling in cell types of interest, we performed tissue-specific CRISPR (*6*) of *BRI1*. We used a *bri1* mutant complemented with *pBRI1-BRI1-mCitrine* (Fig. 4A) into which we introduced *Cas9* driven by tissue-specific promoters to knock out *BRI1* either in the epidermis and lateral root cap (*pWER-BRI1-CRISPR*) or in the cortex (*pCO2-BRI1-CRISPR*). mCitrine signals were absent in the expected locations of the tissue-specific CRISPR lines, confirming their efficacy and specificity (Fig. 4, A to C, and fig. S11, A to C).

Because our scRNA-seq data indicated that brassinosteroids promote the expression of cell wall–related genes in the elongating cortex, we hypothesized that loss of brassinosteroid signaling in the cortex would affect final cell size. Indeed, *pCO2-BRI1-CRISPR* lines displayed significantly shorter mature cortex cells, whereas meristematic cortex cell length was relatively unaffected (Fig. 4, D and E).

By contrast, epidermal knockout of BRI1 in *pWER-BRI1-CRISPR* lines resulted in both reduced meristem cell size and reduced mature cortex cell length (Fig. 4, D and E), which is consistent with the reported role of epidermal brassinosteroid signaling (*9*, *25*, *27*, *35*, *52*) and brassinosteroid responsiveness across developmental zones of the epidermis in our scRNA-seq data.

We measured the cell width of the tissue-specific CRISPR lines and found that *pWER-BRI1-CRISPR* lines have significantly wider cortex cells. However, despite their reduced cortex cell length, *pCO2-BRI1-CRISPR* lines did not have increased cell width (fig. S11D). This prompted us to perform 3D cell segmentations using MorphoGraphX (*55*, *56*). We found that

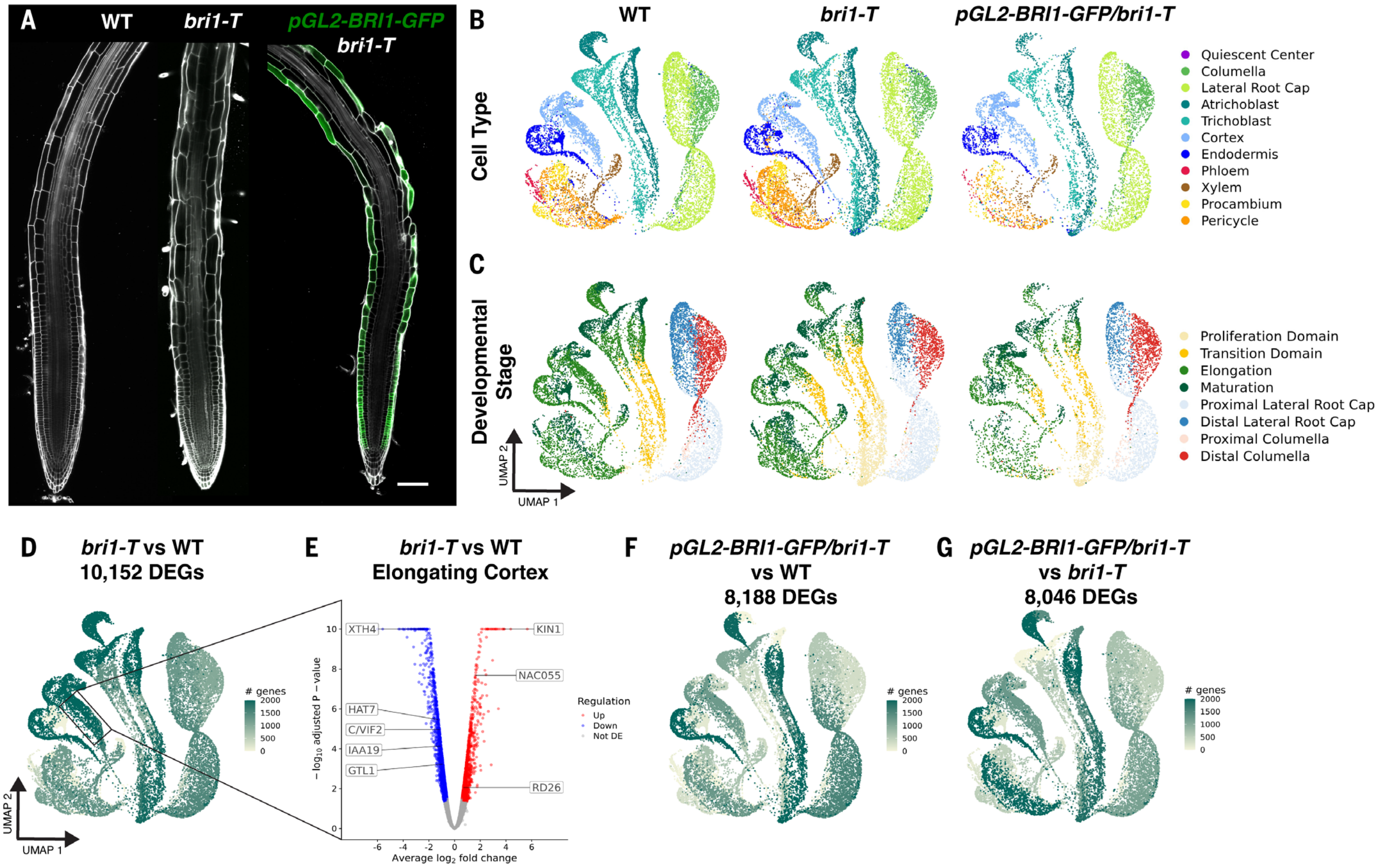


Fig. 3. Triple receptor mutant *bri1-T* gene expression changes in cortex and distinct patterns in *pGL2-BRI1-GFP/bri1-T*. (**A**) Seven-day-old WT, *bri1-T*, and *pGL2-BRI1-GFP/bri1-T* roots grown under control conditions. Propidium iodide staining is shown in gray, and GFP is shown in green. Scale bar, 100 μm. (**B**) UMAP projection of scRNA-seq from 14,334 WT cells, 12,649 *bri1-T* cells, and 7878 *pGL2-BRI1-GFP/bri1-T* cells. Two biological replicates of scRNA-seq were performed for each genotype. Colors indicate cell type annotation. (**C**) UMAP projection colored by developmental stage annotation. (**D**) UMAP colored by DEGs for each cell type–developmental stage combination of *bri1-T* compared with WT. (**E**) Volcano plot of DEGs in the elongating cortex from *bri1-T* compared with WT showing down-regulation of cell wall–related genes in *bri1-T*. Color indicates the direction of regulation. (**F** and **G**) UMAPs colored by DEGs for each cell type–developmental stage combination of *pGL2-BRI1-GFP/bri1-T* compared with WT (F) or *pGL2:BRI1-GFP/bri1-T* compared with *bri1-T* (G).

the cortex cell volume was reduced in *pCO2-BRI1-CRISPR* roots but increased in *pWER-BRI1-CRISPR* roots (fig. S12, A and B). Brassinosteroids primarily control cell elongation rather than cell volume when signaling is affected across the entire root (*51*, *52*). Alternatively, the changes in volume observed in the tissue-specific CRISPR lines could be the result of constraints between tissues or tissue-specific functions of brassinosteroids in influencing cell volume. These results indicate that in addition to the epidermis, brassinosteroid signaling in the cortex is required to promote cell expansion in the elongation zone. The cortex could instruct anisotropic growth through its physical connection with the epidermis, but as the outermost tissue, relaxation of the epidermis appears to be required to allow for cell elongation (*57*, *58*). This may explain the widening of cortex cells in *pWER-BRI1-CRISPR* lines.

BRI1 was also reported to rescue *bri1-T* morphology when expressed in the developing phloem using the *CVP2* promoter (*51*, *53*, *54*). However, gene expression was not fully restored to WT levels in either epidermal or phloem rescue lines. Our scRNA-seq of epidermal *pGL2-BRI1-GFP/bri1-T* lines showed gene expression patterns distinct from either WT or *bri1-T*. Similarly, scRNA-seq of *pCVP2-BRI1-CITRINE/bri1-T* indicated an intermediate state between WT and *bri1-T* (*51*). BRI1 driven by its native promoter was still present in the stele of our tissue-specific CRISPR lines when we observed phenotypic defects, suggesting that, unlike *pCVP2-BRI1*, native expression of *BRI1* in the stele is not sufficient for brassinosteroid-induced cell elongation and root growth. These results confirm the role of the epidermis in brassinosteroid-regulated root growth and reveal the function of the cortex in brassinosteroid-mediated cell elongation, demonstrating how scRNA-seq can identify a spatiotemporal context for hormone signaling.

HAT7 and *GTL1* are brassinosteroid-responsive regulators along cortex trajectories

To define a core set of genes associated with brassinosteroid response along cortex trajectories, we first compared genes induced in the cortex by BL treatment with those down-regulated in the cortex of *bri1-T*. Of the 768 genes in common, we then examined which vary with development in WT cortex trajectories (*31*). The intersection of these three lists identified a core set of 163 brassinosteroid-responsive DEGs (Fig. 5A and data S3). Consistent with regulation by brassinosteroids, 69% of the core DEGs are BES1 and BZR1 direct targets from chromatin immunoprecipitation (ChIP) experiments (*19*, *20*, *59*). Expression along cortex pseudotime illustrates induction by BL treatment and down-regulation in *bri1-T* (Fig. 5B). *HAT7* and *GTL1* were induced along these trajectories, suggesting a potential role for these transcription factors in controlling brassinosteroid-regulated gene expression in the cortex (Fig. 5, C to E).

To gain insight into their roles, we generated translational reporter lines for HAT7 and GTL1. Under control conditions, *pHAT7-HAT7-mCitrine* lines showed expression in the transition domain and elongation zone of the cortex

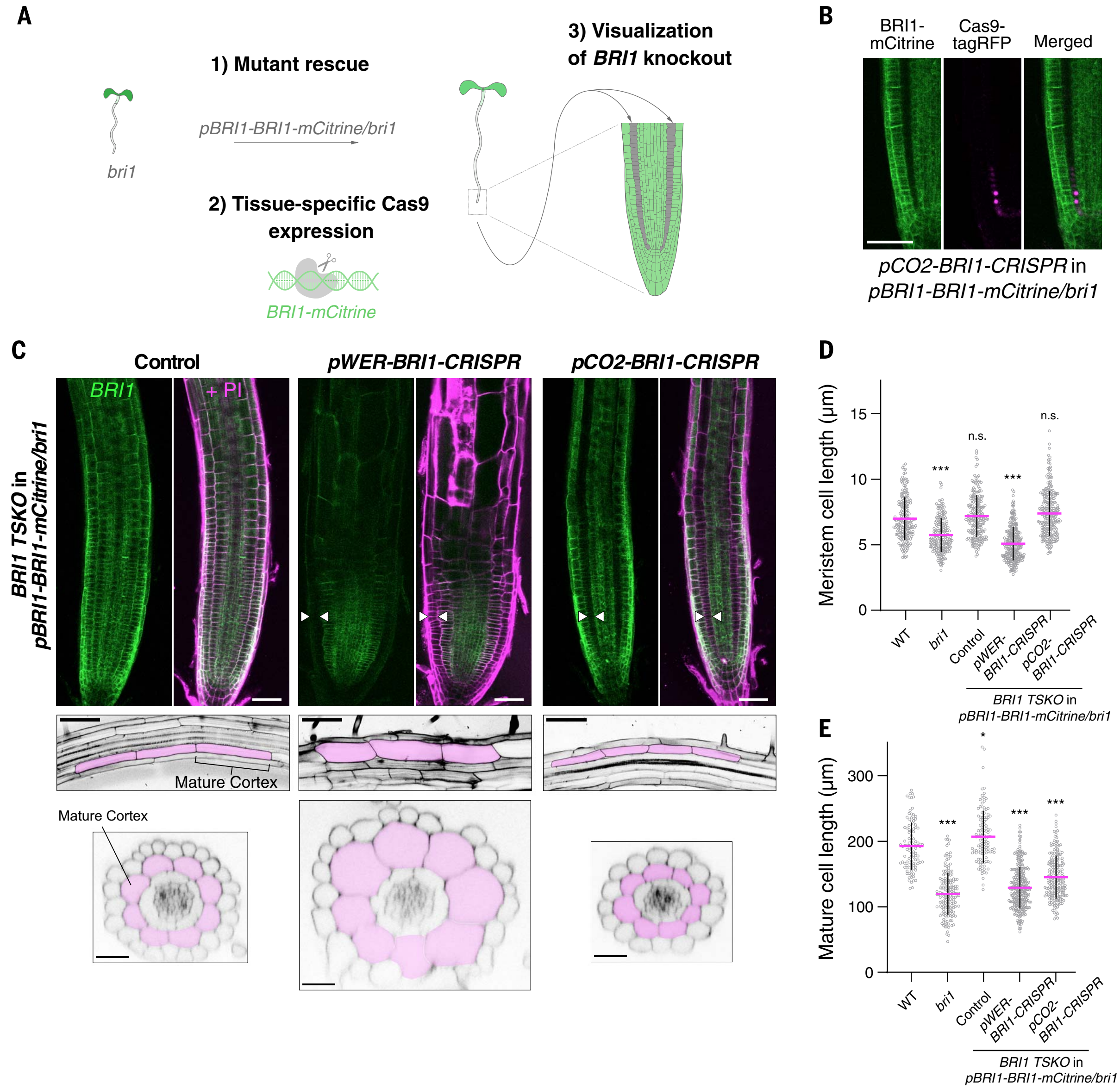


Fig. 4. Tissue-specific CRISPR of *BRI1* confirms role for cortex in brassinosteroid-mediated cell expansion. (**A**) Overview of BRI1 tissue-specific CRISPR approach. A *bri1* mutant complemented with *pBRI1-BRI1-mCitrine* (1) was used as background to introduce tissue-specific Cas9 along with gRNAs targeting BRI1 (2). This allows for visualization of BRI1 knockout in specific cell layers, such as the cortex, when *pCO2-BRI1-CRISPR* is used (3). (**B**) Appearance of Cas9-tagRFP in the cortex is associated with loss of BRI1-mCitrine signal, confirming tissue-specific knockout. RFP, red fluorescent protein. Scale bar, 50 μm. (**C**) Confocal images of BRI1 tissue-specific CRISPR lines. Control indicates a broad expression pattern of BRI1-mCitrine in pBRI1-BRI1-mCitrine/*bri1*. BRI1-mCitrine signals are shown in green and propidium iodide staining (PI) in magenta (top). White arrows specify tissues with absence of BRI1-mCitrine signal; epidermis for *pWER-BRI1-CRISPR* and cortex for *pCO2-BRI1-CRISPR*. Mature root longitudinal and cross sections illustrate changes in cell length (middle) and width (bottom), respectively. Cortex cells are pseudocolored to indicate their position. Scale bars, 50 μm (top), 100 μm (middle), and 25 μm (bottom). (**D**) Quantification of meristematic cortex cell length, defined as the first 20 cells of individual roots starting from the quiescent center. Control indicates pBRI1-BRI1-mCitrine/*bri1* complemented line. (**E**) Quantification of mature cortex cell length. For (D) and (E), all individual data points are plotted. Magenta horizontal bars represent the means, and error bars represent SDs. Significant differences between each line and WT were determined by one-way analysis of variance (ANOVA) and Dunnett's multiple comparison tests. ****P* < 0.001; ***P* < 0.01; **P* < 0.05; n.s., not significant. TSKO, tissue-specific knockout.

(Fig. 5F). We also observed HAT7 signals in the epidermis and endodermis, in line with expression patterns in our WT scRNA-seq atlas (*31*). *HAT7* expression was decreased when brassinosteroid biosynthesis was inhibited with BRZ and restored upon BL treatment (fig. S13A).

pGTL1-GTL1-mCitrine was more broadly expressed, with increasing levels in the cortex as cells progress from the transition domain to the elongation zone (Fig. 5G). GTL1-mCitrine expression was reduced by BRZ and increased by BL treatment (fig. S13B). These results confirm that brassinosteroids promote the expression of *HAT7* and *GTL1*, coinciding with the onset of cell elongation. Furthermore, *HAT7* and *GTL1* are direct targets of BES1 and BZR1 (*19*, *20*, *59*), which suggests that they may be part of the brassinosteroid-directed GRN activated as cells progress from proliferation to elongation.

Previous studies have inferred global (*19*, *20*, *22*, *23*) or temporally resolved GRNs (*21*) for brassinosteroid response, but they have lacked cell type and developmental stage

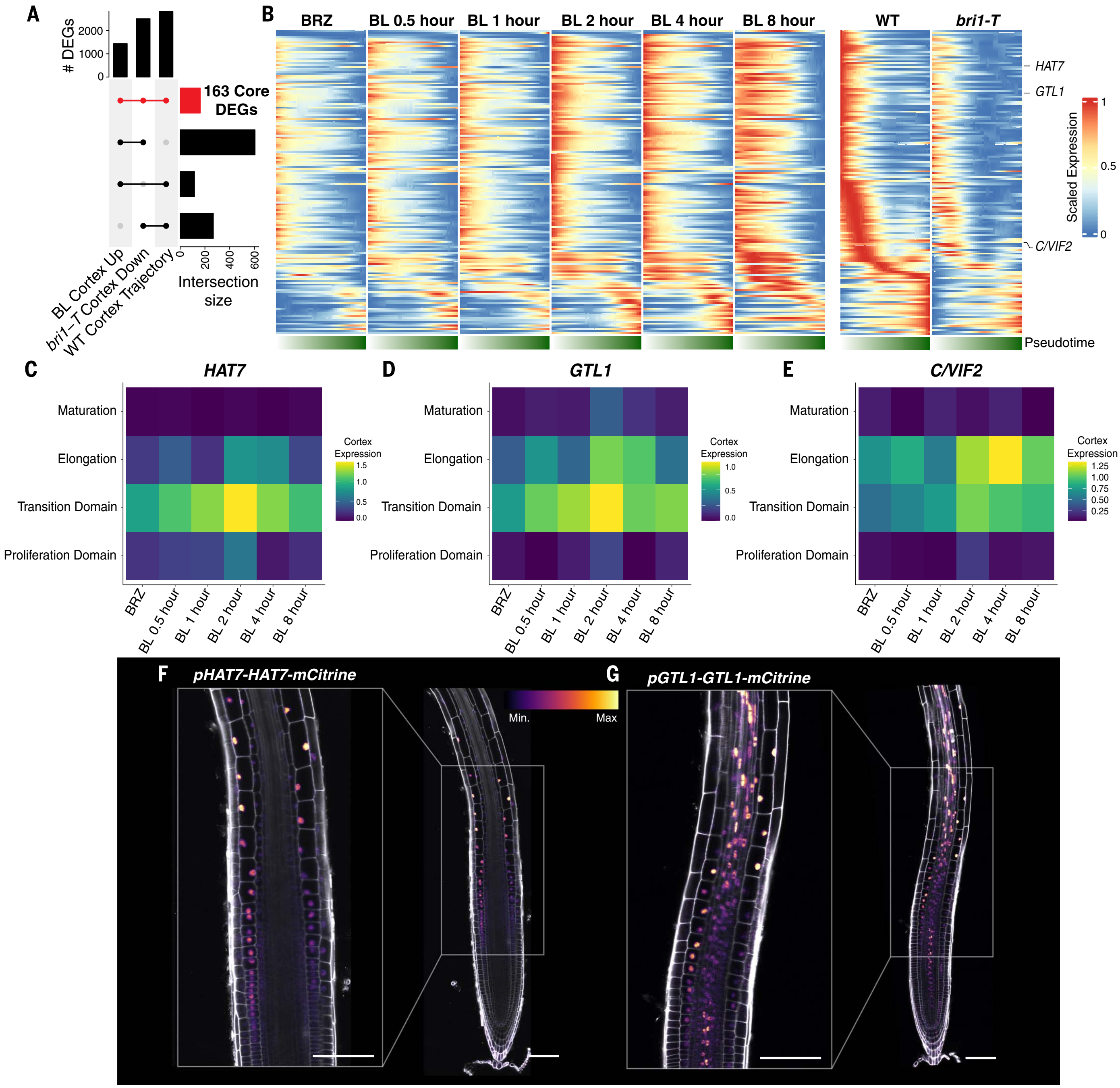


Fig. 5. *HAT7* and *GTL1* are brassinosteroid-responsive regulators along cortex trajectories. (**A**) Upset plot showing a comparison of genes up-regulated by BL in the cortex, down-regulated in the cortex of *bri1-T*, and differentially expressed along WT cortex trajectories. The red color indicates 163 genes common to all three sets. (**B**) Gene expression trends for 163 core brassinosteroid DEGs along cortex trajectories. Scaled expression along cortex pseudotime is plotted for each time point of the brassinosteroid time series and for WT versus *bri1-T*. Lower bar indicates pseudotime progression calculated by CytoTRACE. (**C** to **E**) Gene expression trends for *HAT7* (C), *GTL1* (D), and *C/VIF2* (E) along the developmental zones of the cortex for each time point of the brassinosteroid time course. Color bars indicate the scaled expression level in the cortex. (**F** and **G**) Seven-day-old roots expressing *pHAT7-HAT7-mCitrine* (F) or *pGTL1-GTL1-mCitrine* (G) reporters under control conditions show an increased expression as cortex cells elongate. Propidium iodide staining is shown in gray, with the color gradient indicating relative mCitrine levels. Scale bars, 100 μm.

specificity. To infer GRN configurations across our brassinosteroid time series, we used CellOracle (materials and methods), focusing on BL DEGs and associated transcription factors.

Analysis of network importance scores, such as centrality measures, is a powerful approach to prioritize candidate regulators among DEGs (*60*). Because the cell wall signature peaked at 2 hours after BL treatment, we prioritized transcription factors with high network centrality scores in the elongating cortex at this time point. *HAT7* was the top-ranked transcription factor in terms of degree centrality. *HB13*, *HB20*, and *HB23* were also among the top 10

https://avxhm.se/blogs/hill0

transcription factors (Fig. 6, A and B, and data S4).

We used CRISPR to generate *hat7* loss-of-function mutants but did not observe phenotypes concerning cortex cell elongation (fig. S14, A to C). Together HAT7, HB13, HB20, and HB23 make up the alpha clade HD-ZIP I transcription factors (*45*, *46*, *61–65*). Because *HB13*, *HB20*, and *HB23* are induced by brassinosteroids and are predicted to regulate cell wall–related genes in our GRNs (Fig. 6B; fig. S15, A and B; and data S5), we next generated *hat7 hb13 hb20 hb23* quadruple mutants through multiplex CRISPR. Mature cortex cell length was reduced by ~25% in two independent quadruple mutants (Fig. 6, C and D, and fig. S16, A to C), providing evidence that HAT7 and its homologs are required for cell elongation. *hat7 hb13 hb20 hb23* quadruple mutants also had wider mature cortex cells (fig. S16C) with increased volume compared with WT (fig. S17, A and B). Despite the decrease in final cell length, the root length of the quadruple mutant was not reduced (fig. S14A), which suggests that the decrease in cell length is at least partially compensated for by increased cell production. We found that *hat7 hb13 hb20 hb23* quadruple mutants have an average of 15 more meristematic cortex cells compared with WT (fig. S16B), which is consistent with a compensatory increase in proliferation.

We next investigated GTL1, the fifth-highest-ranked transcription factor in the BL 2-hour elongating cortex GRN (Fig. 6, A and B). Given that GTL1 was shown to function redundantly with DF1 in terminating root hair growth (*66*, *67*), we examined *gtl1 df1* double mutants and found shorter mature cortex cell lengths and shorter roots compared with WT (Fig. 6,

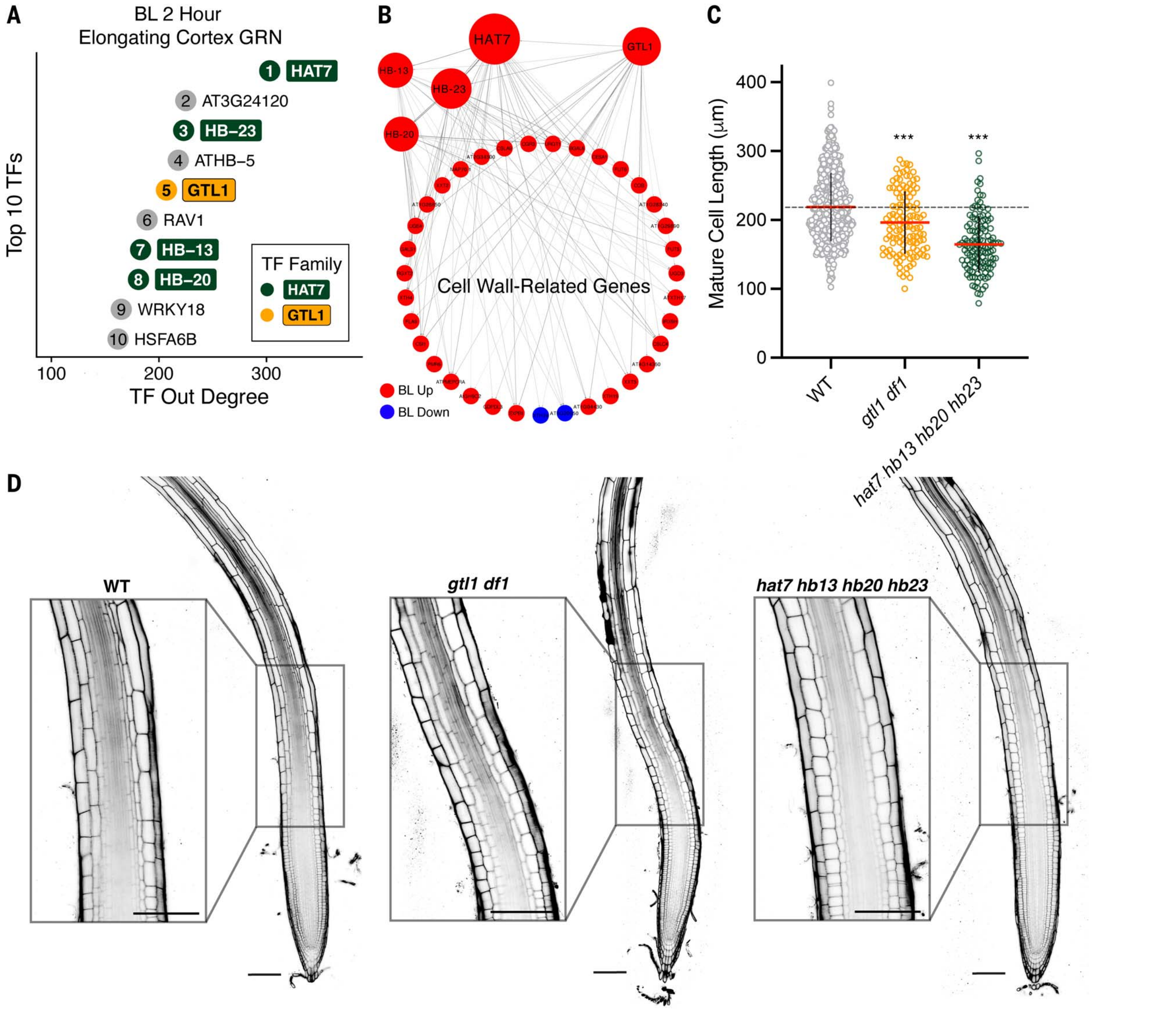


Fig. 6. HAT7 and GTL1 are top-ranked regulators in cortex GRNs and affect brassinosteroid-related phenotypes. (**A**) Top 10 transcription factors (TFs) in the CellOracle BL 2-hour elongating cortex GRN ranked by out-degree. Ranking is indicated by the number inside the circle. Color indicates transcription factor family, with light gray corresponding to any family other than HAT7 or GTL1. (**B**) Subnetwork showing cell wall–related genes that are predicted targets of HAT7 and GTL1 in the CellOracle elongating cortex GRN. HB13, HB20, and HB23 are included in the subnetwork because they are connected to HAT7 and cell wall–related genes. Node size is proportional to degree. (**C**) Quantification of mature cortex cell length. Red horizontal bars represent the means, and error bars represent SDs. Significant differences between each line and WT were determined by one-way ANOVA and Dunnett's multiple comparison tests. $^{***}P < 0.001$. (**D**) Propidium iodide staining of 7-day-old WT, *hat7 hb13 hb20 hb23*, and *gtl1 df1* roots. Insets show cortex cells entering the elongation zone. Scale bars, 100 μm.

C and D, and fig. S14, A to C). *DF1* was challenging to detect in scRNA-seq (fig. S15A) because of its low expression level (*66*). However, we observed increasing trends of *DF1* expression along WOT trajectories in the BL time course, especially in cell wall–responsive cortex cells (fig. S15B), which was verified using a *pDF1-DF1-GFP* reporter (fig. S15C).

We examined the overlap between our CellOracle BL 2-hour elongating cortex GRN and a GRN reported by Clark *et al.* from bulk RNA-seq data in response to brassinosteroids (*21*). Only 43/31,330 edges are shared between the two networks. Moreover, HAT7, HB13, HB20, HB23, and GTL1 were among the top 10 transcription factors in terms of out-degree in the CellOracle elongating cortex GRN, which we validated through mutant analysis. However, none of these were among the top 100 transcription factors in the Clark *et al.* GRN (fig. S18, A and B), which suggests that they would have been difficult to prioritize from the bulk data alone. Together, our genetic analysis of HAT7 and GTL1 family transcription factors illustrates the power of GRN-mediated discovery of regulatory factors in spatiotemporal brassinosteroid response.

BES1 and GTL1 physically interact and regulate shared target genes

Because BES1 is known to interface with other transcription factors in controlling brassinosteroid-regulated gene expression, we compared target genes for BES1 and BZR1 (*19*, *20*, *59*) with ChIP targets of GTL1 and DF1 (*66*). BES1 and BZR1 share 3020 common targets with GTL1 and 2490 common targets with DF1 (fig. S19A). When compared with brassinosteroid-regulated genes from scRNA-seq, BES1 and GTL1 targets showed the strongest enrichment in genes up-regulated by brassinosteroids in the transition domain and elongation zone of the cortex (fig. S19B), with 297 common targets of both BES1 or BZR1 and GTL1 being induced in the elongating cortex by BL treatment.

Given the overlap between BES1 and GTL1 targets, we hypothesized that these transcription factors physically interact to regulate a common set of genes. Coimmunoprecipitation showed that GTL1-FLAG pulled down BES1-GFP (fig. S19C). These results suggest that brassinosteroids induce GTL1 and subsequently BES1 and GTL1 interact to control a common set of target genes. This type of feed-forward loop could provide a mechanism to amplify the brassinosteroid signal and/or to direct BES1 to drive tissue-specific gene expression by interacting with other more specifically expressed transcription factors.

In addition to *GTL1*, we found that previously described BES1 interacting transcription factors, including *BRASSINOSTEROIDS AT VASCULAR AND ORGANIZING CENTER* (*BRAVO*) (*26*), *BIM1* (*42*), and *MYB30* (*68*), were differentially expressed in our scRNA-seq datasets in response to BL (data S3). Among these, *MYB30* was enriched in atrichoblasts of the epidermis and induced by BL treatment (fig. S20, A and B). CellOracle correctly predicted that *MYB30* regulates *LTPG2*, a known target of *MYB30* (*69*), which also showed enrichment in atrichoblasts and temporal induction in response to brassinosteroids after *MYB30* activation (fig. S20B). This supports the idea that cell type–specific patterns of brassinosteroid response may be at least partially explained by interactions between BES1 or BZR1 and other transcription factors.

scRNA-seq reveals cell type–specific expression underlying *gtl1 df1* phenotypes

Our results indicate that *gtl1 df1* mutants have reduced cortex cell elongation. However, *gtl1 df1* mutants have longer trichoblasts (*66*). A downstream regulatory network that enables GTL1-mediated growth inhibition has been dissected in trichoblasts (*66*, *67*). To identify the cell type–specific changes in gene expression underlying *gtl1 df1* cortex phenotypes, we performed scRNA-seq on *gtl1* and *df1* single mutants and the *gtl1 df1* double mutant. Using pseudobulk differential expression analysis, we detected relatively subtle changes in *gtl1* or *df1* single mutants compared with the WT (fig. S21, A to C). By contrast, 8391 genes were differentially expressed in *gtl1 df1* double mutants versus WT (Fig. 7A).

In total, 1077 genes were up-regulated across all developmental stages of the cortex of *gtl1 df1*, and 947 genes were down-regulated. Most cortex DEGs were affected in the elongation zone (Fig. 7, A and B; fig. S21C; and data S3). Of the down-regulated genes in the cortex of the double mutant, 226 genes were also up-regulated by BL treatment. Furthermore, 31.3% of the core brassinosteroid DEGs were down-regulated in the cortex of *gtl1 df1*, whereas only 6.8% were up-regulated (data S3). These results suggest that GTL1 and DF1 promote the expression of a subset of brassinosteroid-induced genes in the cortex.

Plotting *gtl1 df1* DEGs along cortex pseudotime illustrated the down-regulation of several genes involved in cell elongation, including *CESA5* and *AHA2* (Fig. 7C). These genes were enriched for the GO term cell wall organization or biogenesis (fig. S21D). We next examined *C/VIF2* because it is induced by BL in the cortex, but its expression decreased in cortex cells of *gtl1 df1* (Fig. 7, D and E). A *pC/VIF2-H2B-Venus* reporter showed expression of *C/VIF2* in the transition and elongation zone of the WT cortex, whereas its expression was reduced in the cortex of *gtl1 df1* mutants (Fig. 7F and movie S1). The reduced expression of cell wall–related genes in *gtl1 df1* mutants validates our cell type–specific brassinosteroid GRNs and identifies a function of GTL1 in promoting cortex cell elongation in response to brassinosteroids.

Discussion

Understanding how hormone-mediated GRNs are controlled in space and time has the potential to enable the engineering of specific downstream responses to optimize plant growth under a changing environment (*10*, *70*). Plant hormones, including brassinosteroids, auxin, gibberellins, and abscisic acid, have been shown to exhibit tissue-specific responses (*71*–*76*), but how the associated GRNs are modulated in different cell types at particular developmental stages is enigmatic. In this study, we profiled brassinosteroid responses across cell types, developmental stages, and time points of treatment using scRNA-seq, providing a high-resolution map of signaling outputs. These data are publicly available as an interactive browser (https://shiny.mdc-berlin.de/ARVEX/). We identified the elongating cortex as a spatiotemporal context for brassinosteroid signaling, where brassinosteroids activate cell wall–related genes and promote elongation. We further showed that HAT7 and GTL1 are brassinosteroid-induced regulators along cortex trajectories that control cell elongation. These findings highlight the ability of single-cell genomics to identify context-specific transcription factors, a capability that could be leveraged to precisely engineer plant growth, development, and responses to stress. Our results reveal spatiotemporal brassinosteroid responses and the underlying GRNs.

Materials and methods summary

Arabidopsis accession Columbia-0 (Col-0) was used as a WT. The following lines have been previously described: *bri1 GABI_134E10* (*77*); *bri1-116brl1brl3* triple mutant (*bri1-T*) (*50*); *pGL2-BRI1-GFP/bri1-T* (*27*); *gtl1-1* (WiscDsLox413-416C9), *df1-1* (SALK_106258), and *gtl1-1 df1-1* (*66*); and JKD-Ypet recombineering line (*44*). We produced *hat7* single mutants and *hat7 hb13 hb20 hb23* quadruple mutants using FASTRED multiplex CRISPR constructs containing an intronized version of Cas9 (*78*, *79*).

scRNA-seq experiments were performed as previously described (*31*). For each sample, ~0.5-cm root tips from 7-day-old plants were harvested and digested with protoplasting solution, and 16,000 cells were loaded on a 10X Genomics Chromium instrument, with the aim to capture 10,000 cells per sample with the 10X Genomics Chromium 3′ Gene expression v3 or v3.1 kits.

For BL scRNA-seq, we first grew plants on 1 μM BRZ to deplete endogenous brassinosteroids, then transferred plants to either a fresh BRZ plate or 100 nM BL. We performed two separate BL scRNA-seq treatment experiments. The first pilot experiment consisted of a BRZ

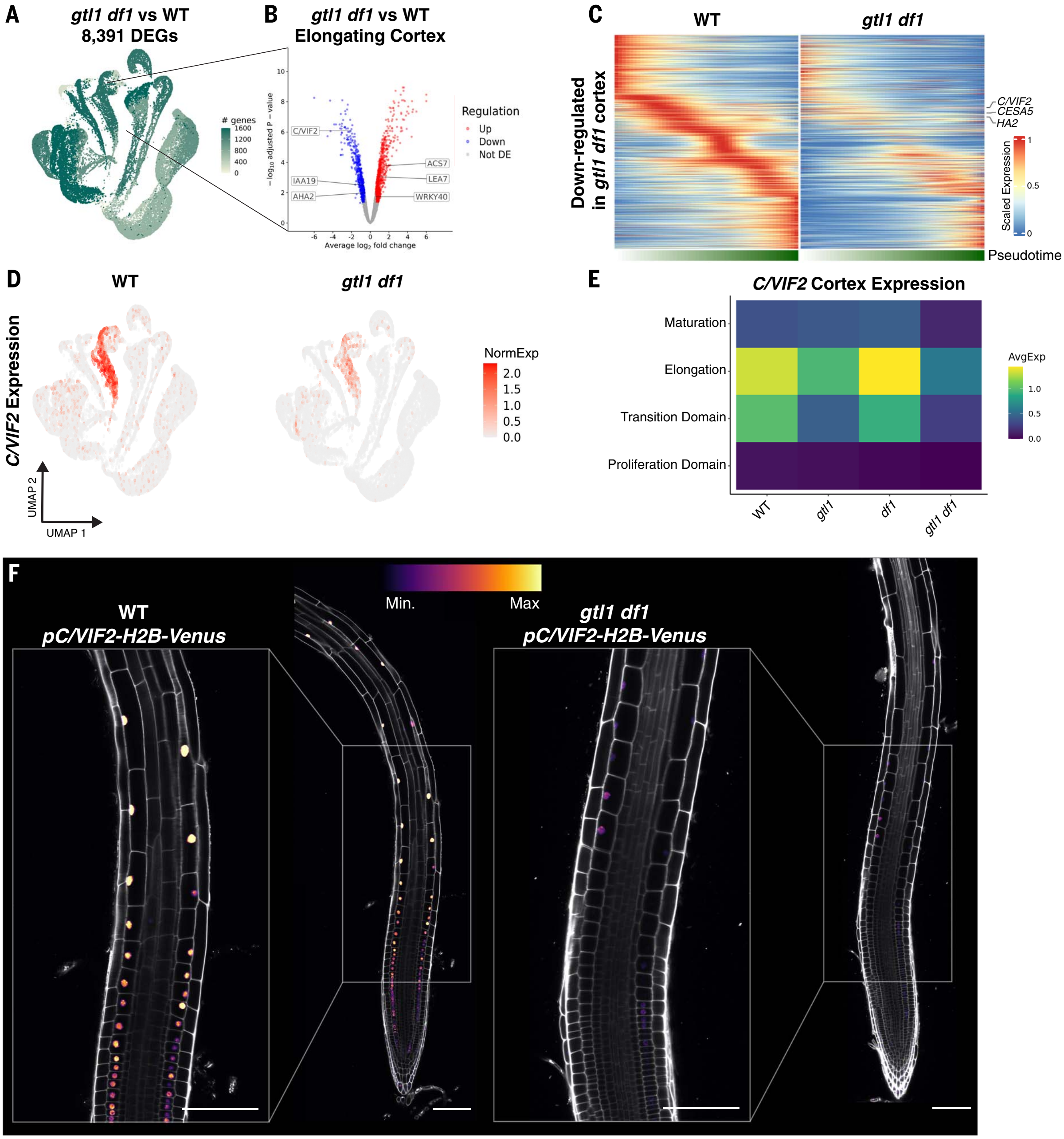


Fig. 7. scRNA-seq reveals cell type–specific expression underlying *gtl1 df1* phenotypes. (**A**) UMAP projection of scRNA-seq from 74,810 WT, *gtl1*, *df1*, and *gtl1 df1* cells. Two biological replicates were profiled for each genotype. Color indicates DEGs for each cell type–developmental stage combination of *gtl1 df1* compared with WT. (**B**) Volcano plot of DEGs in the elongating cortex from *gtl1 df1* compared with WT. Color indicates the direction of regulation. (**C**) Gene expression trends along cortex trajectories for down-regulated DEGs in *gtl1 df1* compared with WT. Each row represents the scaled expression of a gene along cortex pseudotime. The lower bar indicates pseudotime progression calculated by CytoTRACE. (**D**) Expression of *C/VIF2* in WT and *gtl1 df1* scRNA-seq. The color scale represents log normalized, corrected UMI counts. (**E**) *C/VIF2* expression levels plotted along the developmental zones of the cortex for WT, *gtl1*, *df1*, and *gtl1 df1*. The color bar indicates the scaled expression level. (**F**) Seven-day-old root images of a *pC/VIF2-H2B-Venus* reporter in WT or *gtl1 df1* under control conditions. Propidium iodide staining is shown in gray, with the color gradient indicating relative mCitrine levels. Scale bars, 100 μm.

and 2-hour BL treatment. The second experiment included two additional BRZ and BL 2-hour replicates and a single replicate of the other time points in our time course (BL 0.5-, 1-, 4-, and 8-hour treatments). Each of the BL treatments was staggered so that all samples were collected simultaneously. A total of 70,223 cells were recovered from the BL treatment scRNA-seq experiments. WT, *bri1-T*, and *pGL2-BRI1-GFP/bri1-T* were similarly profiled in a side-by-side scRNA-seq experiment under control conditions with two replicates per genotype, resulting in 34,861 cells. To test the effect of inhibiting endogenous brassinosteroid biosynthesis, we grew WT on 1 μM BRZ or a mock dimethyl sulfoxide (DMSO) control and performed two replicates of scRNA-seq spanning 30,962 cells for BRZ versus control analysis.

Finally, scRNA-seq was performed on WT, *gtl1*, *df1*, and *gtl1 df1* in duplicate under control conditions, resulting in 74,810 scRNA-seq expression profiles.

scRNA-seq data were aligned against the *Arabidopsis* TAIR10 reference genome. Quality filtering of cells was performed using the R package COPILOT (Cell preprOcessing PIpeline kaLlistO busTools) (*31*, *80*). Downstream analyses were carried out using Seurat version 3.1.5 (*81*), WOT (*41*), muscat (*82*), tradeSeq (*83*), and GRNs inferred using CellOracle (*60*).

To selectively block brassinosteroid signaling in cell types of interest, we performed tissue-specific CRISPR (*6*) of *BRI1* in a *bri1* mutant complemented with *pBRI1-BRI1-mCitrine*. Two guide RNAs (gRNAs) targeting *BRI1* were simultaneously expressed along with tissue-specific Cas9 to knock out *BRI1* either in the epidermis and lateral root cap (*pWER-BRI1-CRISPR*) or in the cortex (*pCO2-BRI1-CRISPR*). For each root used for quantitative analysis, BRI1-mCitrine signal was acquired to confirm the efficiency of the tissue-specific knockout system. The supplementary materials provide complete details of the materials and methods, including those summarized above.

REFERENCES AND NOTES

1. E. Pierre-Jerome, C. Drapek, P. N. Benfey, Regulation of Division and Differentiation of Plant Stem Cells. *Annu. Rev. Cell Dev. Biol.* **34**, 289–310 (2018). doi: 10.1146/annurev-cellbio-100617-062459; pmid: 30134119
2. R. Shahan, T. M. Nolan, P. N. Benfey, Single-cell analysis of cell identity in the Arabidopsis root apical meristem: Insights and opportunities. *J. Exp. Bot.* **72**, 6679–6686 (2021). doi: 10.1093/jxb/erab228; pmid: 34018001
3. M. Levine, E. H. Davidson, Gene regulatory networks for development. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4936–4942 (2005). doi: 10.1073/pnas.0408031102; pmid: 15788537
4. M. A. Moreno-Risueno, W. Busch, P. N. Benfey, Omics meet networks – using systems approaches to infer regulatory networks in plants. *Curr. Opin. Plant Biol.* **13**, 126–131 (2010). doi: 10.1016/j.pbi.2009.11.005; pmid: 20036612
5. C. Seyfferth *et al*., Advances and Opportunities in Single-Cell Transcriptomics for Plant Research. *Annu. Rev. Plant Biol.* **72**, 847–866 (2021). doi: 10.1146/annurev-arplant-081720-010120; pmid: 33730513
6. W. Decaestecker *et al*., CRISPR-TSKO: A Technique for Efficient Mutagenesis in Specific Cell Types, Tissues, or Organs in Arabidopsis. *Plant Cell* **31**, 2868–2887 (2019). doi: 10.1105/tpc.19.00454; pmid: 31562216
7. X. Wang *et al*., An inducible genome editing system for plants. *Nat. Plants* **6**, 766–772 (2020). doi: 10.1038/s41477-020-0695-2; pmid: 32601420
8. S. D. Clouse, M. Langford, T. C. McMorris, A brassinosteroid-insensitive mutant in Arabidopsis thaliana exhibits multiple defects in growth and development. *Plant Physiol.* **111**, 671–678 (1996). doi: 10.1104/pp.111.3.671; pmid: 8754677
9. Y. Hacham *et al*., Brassinosteroid perception in the epidermis controls root meristem size. *Development* **138**, 839–848 (2011). doi: 10.1242/dev.061804; pmid: 21270053
10. T. M. Nolan, N. Vukašinović, D. Liu, E. Russinova, Y. Yin, Brassinosteroids: Multidimensional Regulators of Plant Growth, Development, and Stress Responses. *Plant Cell* **32**, 295–318 (2020). doi: 10.1105/tpc.19.00335; pmid: 31776234
11. J. Li, P. Nagpal, V. Vitart, T. C. McMorris, J. Chory, A role for brassinosteroids in light-dependent development of *Arabidopsis*. *Science* **272**, 398–401 (1996). doi: 10.1126/science.272.5260.398; pmid: 8602526
12. M.-P. González-García *et al*., Brassinosteroids control meristem size by promoting cell cycle progression in *Arabidopsis* roots. *Development* **138**, 849–859 (2011). doi: 10.1242/dev.057331; pmid: 21270057
13. A. Caño-Delgado *et al*., BRL1 and BRL3 are novel brassinosteroid receptors that function in vascular differentiation in *Arabidopsis*. *Development* **131**, 5341–5351 (2004). doi: 10.1242/dev.01403; pmid: 15486337
14. T. Kinoshita *et al*., Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* **433**, 167–171 (2005). doi: 10.1038/nature03227; pmid: 15650741
15. J. Li, J. Chory, A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* **90**, 929–938 (1997). doi: 10.1016/S0092-8674(00)80357-8; pmid: 9298904
16. J. Li *et al*., BAK1, an *Arabidopsis* LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell* **110**, 213–222 (2002). doi: 10.1016/S0092-8674(02)00812-7; pmid: 12150929
17. Y. Yin *et al*., BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation. *Cell* **109**, 181–191 (2002). doi: 10.1016/S0092-8674(02)00721-3; pmid: 12007405
18. Z. Y. Wang *et al*., Nuclear-localized BZR1 mediates brassinosteroid-induced growth and feedback suppression of brassinosteroid biosynthesis. *Dev. Cell* **2**, 505–513 (2002). doi: 10.1016/S1534-5807(02)00153-3; pmid: 11970900
19. Y. Sun *et al*., Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in *Arabidopsis*. *Dev. Cell* **19**, 765–777 (2010). doi: 10.1016/j.devcel.2010.10.010; pmid: 21074725
20. X. Yu *et al*., A brassinosteroid transcriptional network revealed by genome-wide identification of BESI target genes in *Arabidopsis thaliana*. *Plant J.* **65**, 634–646 (2011). doi: 10.1111/j.1365-313X.2010.04449.x; pmid: 21214652
21. N. M. Clark *et al*., Integrated omics networks reveal the temporal signaling events of brassinosteroid response in *Arabidopsis*. *Nat. Commun.* **12**, 5858 (2021). doi: 10.1038/s41467-021-26165-3; pmid: 34615886
22. R. Seyed Rahmani *et al*., Genome-wide expression and network analyses of mutants in key brassinosteroid signaling genes. *BMC Genomics* **22**, 465 (2021). doi: 10.1186/s12864-021-07778-w; pmid: 34157989
23. H. Guo, L. Li, M. Aluru, S. Aluru, Y. Yin, Mechanisms and networks for brassinosteroid regulated gene expression. *Curr. Opin. Plant Biol.* **16**, 545–553 (2013). doi: 10.1016/j.pbi.2013.08.002; pmid: 23993372
24. Y. Fridman *et al*., Root growth is modulated by differential hormonal sensitivity in neighboring cells. *Genes Dev.* **28**, 912–920 (2014). doi: 10.1101/gad.239335.114; pmid: 24736847
25. M. Ackerman-Lavert *et al*., Auxin requirements for a meristematic state in roots depend on a dual brassinosteroid function. *Curr. Biol.* **31**, 4462–4472.e6 (2021). doi: 10.1016/j.cub.2021.07.075; pmid: 34418341
26. J. Vilarrasa-Blasi *et al*., Regulation of plant stem cell quiescence by a brassinosteroid signaling module. *Dev. Cell* **30**, 36–47 (2014). doi: 10.1016/j.devcel.2014.05.020; pmid: 24981610
27. K. Vragović *et al*., Translatome analyses capture of opposing tissue-specific brassinosteroid signals orchestrating root meristem differentiation. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 923–928 (2015). doi: 10.1073/pnas.1417947112; pmid: 25561530
28. I. Betegón-Putze *et al*., Precise transcriptional control of cellular quiescence by BRAVO/WOX5 complex in *Arabidopsis* roots. *Mol. Syst. Biol.* **17**, e9864 (2021). doi: 10.15252/msb.20209864; pmid: 34132490
29. T. Asami *et al*., Characterization of brassinazole, a triazole-type brassinosteroid biosynthesis inhibitor. *Plant Physiol.* **123**, 93–100 (2000). doi: 10.1104/pp.123.1.93; pmid: 10806228
30. M. D. Grove *et al*., Brassinolide, a plant growth-promoting steroid isolated from *Brassica napus* pollen. *Nature* **281**, 216–217 (1979). doi: 10.1038/281216a0
31. R. Shahan *et al*., A single-cell *Arabidopsis* root atlas reveals developmental trajectories in wild-type and cell identity mutants. *Dev. Cell* **57**, 543–560.e9 (2022). doi: 10.1016/j.devcel.2022.01.008; pmid: 35134336
32. V. B. Ivanov, J. G. Dubrovsky, Longitudinal zonation pattern in plant roots: Conflicts and solutions. *Trends Plant Sci.* **18**, 237–243 (2013). doi: 10.1016/j.tplants.2012.10.002; pmid: 23123304
33. E. Salvi, R. Di Mambro, S. Sabatini, Dissecting mechanisms in root growth from the transition zone perspective. *J. Exp. Bot.* **71**, 2390–2396 (2020). doi: 10.1093/jxb/eraa079; pmid: 32064533
34. T. Denyer *et al*., Spatiotemporal Developmental Trajectories in the *Arabidopsis* Root Revealed Using High-Throughput Single-Cell RNA Sequencing. *Dev. Cell* **48**, 840–852.e5 (2019). doi: 10.1016/j.devcel.2019.02.022; pmid: 30913408
35. J. Chaiwanon, Z.-Y. Wang, Spatiotemporal brassinosteroid signaling and antagonism with auxin pattern stem cell dynamics in *Arabidopsis* roots. *Curr. Biol.* **25**, 1031–1042 (2015). doi: 10.1016/j.cub.2015.02.046; pmid: 25866388
36. P. E. Verslues, T. Longkumer, Size and activity of the root meristem: A key for drought resistance and a key model of drought-related signaling. *Physiol. Plant.* **174**, e13622 (2022). doi: 10.1111/ppl.13622; pmid: 34988997
37. T. Longkumer, C.-Y. Chen, M. Biancucci, G. B. Bhaskara, P. E. Verslues, Spatial differences in stoichiometry of EGR phosphatase and Microtubule-associated Stress Protein 1 control root meristem activity during drought stress. *Plant Cell* **34**, 742–758 (2022). doi: 10.1093/plcell/koab290; pmid: 34865106
38. D. Dietrich *et al*., Root hydrotropism is controlled via a cortex-specific growth mechanism. *Nat. Plants* **3**, 17057 (2017). doi: 10.1038/nplants.2017.57; pmid: 28481327
39. R. Miao *et al*., Low ABA concentration promotes root growth and hydrotropism through relief of ABA INSENSITIVE 1-mediated inhibition of plasma membrane H^+-ATPase 2. *Sci. Adv.* **7**, eabd4113 (2021). doi: 10.1126/sciadv.abd4113; pmid: 33731345
40. L. Xie, C. Yang, X. Wang, Brassinosteroids can regulate cellulose biosynthesis by controlling the expression of CESA genes in *Arabidopsis*. *J. Exp. Bot.* **62**, 4495–4506 (2011). doi: 10.1093/jxb/err164; pmid: 21617247
41. G. Schiebinger *et al*., Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 928–943.e22 (2019). doi: 10.1016/j.cell.2019.02.026; pmid: 30849376
42. Y. Yin *et al*., A new class of transcription factors mediates brassinosteroid-regulated gene expression in *Arabidopsis*. *Cell* **120**, 249–259 (2005). doi: 10.1016/j.cell.2004.11.044; pmid: 15680330
43. M. K. Zhiponova *et al*., Helix-loop-helix/basic helix-loop-helix transcription factor network represses cell elongation in *Arabidopsis* through an apparent incoherent feed-forward loop. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2824–2829 (2014). doi: 10.1073/pnas.1400203111; pmid: 24505057
44. M. A. Moreno-Risueno *et al*., Transcriptional control of tissue formation throughout root development. *Science* **350**, 426–430 (2015). doi: 10.1126/science.aad1171; pmid: 26494755
45. E. Henriksson *et al*., Homeodomain leucine zipper class I genes in Arabidopsis. Expression patterns and phylogenetic relationships. *Plant Physiol.* **139**, 509–518 (2005). doi: 10.1104/pp.105.063461; pmid: 16055682
46. J. Mattsson, E. Söderman, M. Svenson, C. Borkird, P. Engström, A new homeobox-leucine zipper gene from *Arabidopsis thaliana*. *Plant Mol. Biol.* **18**, 1019–1022 (1992). doi: 10.1007/BF00019223; pmid: 1349836
47. N. Vukašinović *et al*., Local brassinosteroid biosynthesis enables optimal root growth. *Nat. Plants* **7**, 619–632 (2021). doi: 10.1038/s41477-021-00917-x; pmid: 34007032
48. D. M. Friedrichsen, C. A. Joazeiro, J. Li, T. Hunter, J. Chory, Brassinosteroid-insensitive-1 is a ubiquitously expressed leucine-rich repeat receptor serine/threonine kinase. *Plant Physiol.* **123**, 1247–1256 (2000). doi: 10.1104/pp.123.4.1247; pmid: 10938344
49. Z. He *et al*., Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. *Science* **288**, 2360–2363 (2000). doi: 10.1126/science.288.5475.2360; pmid: 10875920
50. N. G. Irani *et al*., Fluorescent castasterone reveals BRI1 signaling from the plasma membrane. *Nat. Chem. Biol.* **8**, 583–589 (2012). doi: 10.1038/nchembio.958; pmid: 22561410
51. M. Graeff *et al*., A single-cell morpho-transcriptomic map of brassinosteroid action in the *Arabidopsis* root. *Mol. Plant* **14**, 1985–1999 (2021). doi: 10.1016/j.molp.2021.07.021; pmid: 34358681
52. Y. Fridman *et al*., The root meristem is shaped by brassinosteroid control of cell geometry. *Nat. Plants* **7**, 1475–1484 (2021). doi: 10.1038/s41477-021-01014-9; pmid: 34782771
53. Y. H. Kang, A. Breda, C. S. Hardtke, Brassinosteroid signaling directs formative cell divisions and protophloem differentiation in *Arabidopsis* root meristems. *Development* **144**, 272–280 (2017). doi: 10.1242/dev.145623; pmid: 28096215
54. M. Graeff, S. Rana, P. Marhava, B. Moret, C. S. Hardtke, Local and Systemic Effects of Brassinosteroid Perception in

Developing Phloem. *Curr. Biol.* **30**, 1626–1638.e3 (2020). doi: 10.1016/j.cub.2020.02.029; pmid: 32220322
55. P. Barbier de Reuille *et al*., MorphoGraphX: A platform for quantifying morphogenesis in 4D. *eLife* **4**, e05864 (2015). doi: 10.7554/eLife.05864; pmid: 25946108
56. S. Strauss *et al*., Using positional information to provide context for biological image analysis with MorphoGraphX 2.0. *eLife* **11**, e72601 (2022). doi: 10.7554/eLife.72601; pmid: 35510843
57. F. Bou Daher *et al*., Anisotropic growth is achieved through the additive mechanical effect of material anisotropy and elastic asymmetry. *eLife* **7**, e38161 (2018). doi: 10.7554/eLife.38161; pmid: 30226465
58. T. I. Baskin, O. E. Jensen, On the role of stress anisotropy in the growth of stems. *J. Exp. Bot.* **64**, 4697–4707 (2013). doi: 10.1093/jxb/ert176; pmid: 23913952
59. E. Oh *et al*., Cell elongation is regulated through a central circuit of interacting transcription factors in the Arabidopsis hypocotyl. *eLife* **3**, e03031 (2014). doi: 10.7554/eLife.03031; pmid: 24867218
60. K. Kamimoto *et al*., Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742–751 (2023). doi: 10.1038/s41586-022-05688-9; pmid: 36755098
61. F. D. Ariel, P. A. Manavella, C. A. Dezar, R. L. Chan, The true story of the HD-Zip family. *Trends Plant Sci.* **12**, 419–426 (2007). doi: 10.1016/j.tplants.2007.08.003; pmid: 17698401
62. M. F. Perotti, P. A. Ribone, J. V. Cabello, F. D. Ariel, R. L. Chan, AtHB23 participates in the gene regulatory network controlling root branching, and reveals differences between secondary and tertiary roots. *Plant J.* **100**, 1224–1236 (2019). doi: 10.1111/tpj.14511; pmid: 31444832
63. A. T. Silva, P. A. Ribone, R. L. Chan, W. Ligterink, H. W. M. Hilhorst, A Predictive Coexpression Network Identifies Novel Genes Controlling the Seed-to-Seedling Phase Transition in *Arabidopsis thaliana*. *Plant Physiol.* **170**, 2218–2231 (2016). doi: 10.1104/pp.15.01704; pmid: 26888061
64. P. A. Ribone, M. Capella, R. L. Chan, Functional characterization of the homeodomain leucine zipper I transcription factor AtHB13 reveals a crucial role in *Arabidopsis* development. *J. Exp. Bot.* **66**, 5929–5943 (2015). doi: 10.1093/jxb/erv302; pmid: 26136262
65. J. C. Harris, M. Hrmova, S. Lopato, P. Langridge, Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli. *New Phytol.* **190**, 823–837 (2011). doi: 10.1111/j.1469-8137.2011.03733.x; pmid: 21517872
66. M. Shibata *et al*., GTL1 and DF1 regulate root hair growth through transcriptional repression of *ROOT HAIR DEFECTIVE 6-LIKE 4* in *Arabidopsis*. *Development* **145**, dev159707 (2018). doi: 10.1242/dev.159707; pmid: 29439132
67. M. Shibata *et al*., GTL1 is required for a robust root hair growth response to avoid nutrient overloading. *bioRxiv* 2021.06.27.450011 [Preprint] (2021). doi: 10.1101/2021.06.27.450011
68. L. Li *et al*., Arabidopsis MYB30 is a direct target of BES1 and cooperates with BES1 to regulate brassinosteroid-induced gene expression. *Plant J.* **58**, 275–286 (2009). doi: 10.1111/j.1365-313X.2008.03778.x; pmid: 19170933
69. K. Mabuchi *et al*., MYB30 links ROS signaling, root cell elongation, and plant immune responses. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4710–E4719 (2018). doi: 10.1126/science.aaz7614; pmid: 32299946
70. A. Gupta, A. Rico-Medina, A. I. Caño-Delgado, The physiology of plant responses to drought. *Science* **368**, 266–269 (2020). doi: 10.1126/science.aaz7614; pmid: 32299946
71. B. O. R. Bargmann *et al*., A map of cell type-specific auxin responses. *Mol. Syst. Biol.* **9**, 688 (2013). doi: 10.1038/msb.2013.40; pmid: 24022006
72. E. Shani *et al*., Gibberellins accumulate in the elongating endodermal cells of *Arabidopsis* root. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4834–4839 (2013). doi: 10.1073/pnas.1300436110; pmid: 23382232
73. S. Ubeda-Tomás *et al*., Root growth in *Arabidopsis* requires gibberellin/DELLA signalling in the endodermis. *Nat. Cell Biol.* **10**, 625–628 (2008). doi: 10.1038/ncb1726; pmid: 18425113
74. Y. Geng *et al*., A spatio-temporal understanding of growth regulation during the salt stress response in *Arabidopsis*. *Plant Cell* **25**, 2132–2154 (2013). doi: 10.1105/tpc.113.112896; pmid: 23898029
75. A. S. Iyer-Pascuzzi *et al*., Cell identity regulators link development and stress responses in the *Arabidopsis* root. *Dev. Cell* **21**, 770–782 (2011). doi: 10.1016/j.devcel.2011.09.009; pmid: 22014526
76. M. Ackerman-Lavert, S. Savaldi-Goldstein, Growth models from a brassinosteroid perspective. *Curr. Opin. Plant Biol.* **53**, 90–97 (2020). doi: 10.1016/j.pbi.2019.10.008; pmid: 31809963
77. Y. Jaillais, Y. Belkhadir, E. Balsemão-Pires, J. L. Dangl, J. Chory, Extracellular leucine-rich repeats as a platform for receptor/coreceptor complex formation. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8503–8507 (2011). doi: 10.1073/pnas.1103556108; pmid: 21464298
78. J. Stuttmann *et al*., Highly efficient multiplex editing: One-shot generation of 8× *Nicotiana benthamiana* and 12× Arabidopsis mutants. *Plant J.* **106**, 8–22 (2021). doi: 10.1111/tpj.15197; pmid: 33577114
79. R. Grützner *et al*., High-efficiency genome editing in plants mediated by a Cas9 gene containing multiple introns. *Plant Commun.* **2**, 100135 (2020). doi: 10.1016/j.xplc.2020.100135; pmid: 33898975
80. C.-W. Hsu, R. Shahan, T. M. Nolan, P. N. Benfey, U. Ohler, Protocol for fast scRNA-seq raw data processing using scKB and non-arbitrary quality control with COPILOT. *STAR Protoc.* **3**, 101729 (2022). doi: 10.1016/j.xpro.2022.101729; pmid: 36181683
81. T. Stuart *et al*., Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019). doi: 10.1016/j.cell.2019.05.031; pmid: 31178118
82. H. L. Crowell *et al*., *muscat* detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nat. Commun.* **11**, 6077 (2020). doi: 10.1038/s41467-020-19894-4; pmid: 33257685
83. K. Van den Berge *et al*., Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1201 (2020). doi: 10.1038/s41467-020-14766-3; pmid: 32139671
84. T. Nolan, C.-W. Hsu, L. Greenstreet, tmnolan/Brassinosteroid-gene-regulatory-networks-at-cellular-resolution: Brassinosteroid gene regulatory networks at cellular resolution in the Arabidopsis root, version 1.0.0, Zenodo (2022); doi: 10.5281/zenodo.7489359
85. T. Nolan, Brassinosteroid gene regulatory networks at cellular resolution in the Arabidopsis root, data set, *Zenodo* (2022); doi: 10.5281/zenodo.7489389

ACKNOWLEDGMENTS

The authors thank H. Belcher and M. Perkins Jacobs for technical assistance; N. Devos and Duke GCB for sequencing services; N. Vangheluwe, T. Beeckman, and T. B. Jacobs (VIB-UGent) for CRISPR-TSKO advice; and K. Sugimoto for *gtl1* and *df1* seeds. **Funding:** This study was supported by the Howard Hughes Medical Institute (P.N.B.); National Institutes of Health MIRA 1R35GM131725 (P.N.B.); National Science Foundation Postdoctoral Research Fellowships in Biology Program grant IOS-2010686 (T.M.N.); National Institutes of Health NRSA postdoctoral fellowship 1F32GM136030-01 (R.S.); Research Foundation-Flanders G002121N (E.R.); Research Foundation-Flanders Postdoctoral fellowships nos. 12R7822N and 12R7819N (N.V.); US Department of Agriculture National Institute of Food and Agriculture fellowship 2021-67034-35139 (I.W.T.); National Science Foundation grant MCB 1818160 (Y.Y.); Deutsche Forschungsgemeinschaft International Research Training Group 2403 (C.-W.H. and U.O.); a Career Award at the Scientific Interface from the Burroughs Wellcome Fund (L.G., M.H., and G.S.); a New Frontiers in Research Fund (NFRF) exploration grant (L.G., M.H., A.A., and G.S.); a Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant (L.G., M.H., A.A., and G.S.); and a project grant from the Canadian Institute of Health Research (L.G., M.H., A.A., and G.S.). **Author contributions:** Conceptualization: T.M.N., N.V., C.-W.H., E.R., and P.N.B. Methodology: T.M.N., N.V., C.-W.H., J.Z., P.S., I.V., A.B., and P.W. Investigation: T.M.N., N.V., R.S., I.W.T., J.Z., P.S., I.V., A.B., and P.W. Data analysis: C.-W.H., T.M.N., L.G., M.H., A.A., and G.S. Writing: T.M.N., C.-W.H., N.V., R.S., I.W.T., E.R., and P.N.B. Supervision: P.N.B., E.R., U.O., G.S., and Y.Y. **Competing interests:** P.N.B. is the cofounder and chair of the Scientific Advisory Board of Hi Fidelity Technologies, a company that works on crop root growth. The authors declare no other competing interests. **Data and materials availability:** scRNA-seq data have been deposited at Gene Expression Omnibus (GEO) (accession no. GSE212230) and can be visualized through an interactive browser at https://shiny.mdc-berlin.de/ARVEX/. All data are available in the manuscript or the supplementary materials or are deposited at Zenodo (*84*, *85*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf4721
Materials and Methods
Figs. S1 to S21
References (*86*–*130*)
MDAR Reproducibility Checklist
Movie S1
Data S1 to S6

View/request a protocol for this paper from *Bio-protocol*.

Submitted 24 October 2022; accepted 9 February 2023
10.1126/science.adf4721

RESEARCH ARTICLE

PHYLOGEOGRAPHY

Early dispersal of domestic horses into the Great Plains and northern Rockies

William Timothy Treal Taylor[1,2]*†, Pablo Librado[3]†, Mila Hunska Tašunke Icu (Chief Joseph American Horse)[4]†, Carlton Shield Chief Gover[1,5]†, Jimmy Arterberry[6]†, Anpetu Luta Wiŋ (Antonia Loretta Afraid of Bear-Cook)[4], Akil Nujipi (Harold Left Heron)[7], Tanka Omniya (Robert Milo Yellow Hair)[4], Mario Gonzalez (Nantan Hinapan)[4], Bill Means[4,8], Sam High Crane (Wapageya Mani)[9]‡, Mažasu (Wendell W. Yellow Bull)[4], Barbara Dull Knife (Mah'piya Keyaké Wiŋ)[4,10], Wakiŋyala Wiŋ (Anita Afraid of Bear)[4], Cruz Tecumseh Collin (Wanka'tuya Kiya)[4], Chance Ward[2,11], Theresa A. Pasqual[12], Lorelei Chauvey[3], Laure Tonasso-Calviere[3], Stéphanie Schiavinato[3], Andaine Seguin-Orlando[3], Antoine Fages[3,13], Naveed Khan[3,14], Clio Der Sarkissian[3], Xuexue Liu[3], Stefanie Wagner[3], Beth Ginondidoy Leonard[15,16], Bruce L. Manzano[17], Nancy O'Malley[18], Jennifer A. Leonard[19], Eloísa Bernáldez-Sánchez[20], Eric Barrey[21], Léa Charliquart[22], Emilie Robbe[23], Thibault Denoblet[22], Kristian Gregersen[23], Alisa O. Vershinina[24], Jaco Weinstock[25], Petra Rajić Šikanjić[26], Marjan Mashkour[27], Irina Shingiray[28,29], Jean-Marc Aury[30], Aude Perdereau[30], Saleh Alquraishi[31,32], Ahmed H. Alfarhan[33,34], Khaled A. S. Al-Rasheid[32], Tajana Trbojević Vukičević[35], Marcel Buric[36], Eberhard Sauer[37], Mary Lucas[38], Joan Brenner-Coltrain[39], John R. Bozell[40], Cassidee A. Thornhill[41], Victoria Monagle[42], Angela Perri[43], Cody Newton[44], W. Eugene Hall[45], Joshua L. Conver[46], Petrus Le Roux[47], Sasha G. Buckser[1], Caroline Gabe[48], Juan Bautista Belardi[49], Christina I. Barrón-Ortiz[50], Isaac A. Hart[39], Christina Ryder[1], Matthew Sponheimer[1], Beth Shapiro[51], John Southon[52], Joss Hibbs[53], Charlotte Faulkner[53], Alan Outram[54], Laura Patterson Rosa[55], Katelyn Palermo[56], Marina Solé[57], Alice William[58], Wayne McCrory[59], Gabriella Lindgren[57,60], Samantha Brooks[61], Camille Eché[62], Cécile Donnadieu[62], Olivier Bouchez[62], Patrick Wincker[30], Gregory Hodgins[63], Sarah Trabert[64], Brandi Bethke[65], Patrick Roberts[38,66], Emily Lena Jones[42]†, Yvette Running Horse Collin (Tašunke Iyanke Wiŋ)[3,4]†, Ludovic Orlando[3]†*

The horse is central to many Indigenous cultures across the American Southwest and the Great Plains. However, when and how horses were first integrated into Indigenous lifeways remain contentious, with extant models derived largely from colonial records. We conducted an interdisciplinary study of an assemblage of historic archaeological horse remains, integrating genomic, isotopic, radiocarbon, and paleopathological evidence. Archaeological and modern North American horses show strong Iberian genetic affinities, with later influx from British sources, but no Viking proximity. Horses rapidly spread from the south into the northern Rockies and central plains by the first half of the 17th century CE, likely through Indigenous exchange networks. They were deeply integrated into Indigenous societies before the arrival of 18th-century European observers, as reflected in herd management, ceremonial practices, and culture.

The spread of domestic horses and their integration into Indigenous societies contributed to profound social and ecological transformations across western North America. However, the mechanisms and timing of this transition are poorly understood. Horses and other members of the genus *Equus* originated in North America (*1*, *2*). Horses and equids formed an important component of human lifeways across the continent during the final Pleistocene (*3*–*5*), which is still encoded in some Indigenous oral traditions, including those of the Lakota (*6*). Although Western scholars commonly consider horses to have disappeared at lower latitudes by the early Holocene, environmental DNA suggests their presence in arctic zones as late as 5000 to 6000 years before the present (*7*, *8*). Few archaeozoological studies have carefully addressed their possible persistence at lower latitudes during the Holocene.

Viking colonizers brought horses as far as Greenland during the 10th to 14th centuries CE (*9*) and settled along areas of the Newfoundland coast during the 11th century CE (*10*). There is, however, no direct evidence that Viking horses reached settlements on the mainland (*11*). Instead, most western scholars accept that horses were first reintroduced into the Americas by Spanish settlers in the late 15th century CE, reaching the mainland in the early 16th century CE with the Spanish colonization of Mexico (*12*). During the 17th to 19th centuries CE, colonizing European powers, including the British, Spanish, and French (*13*, *14*), and possibly Russian and Chinese merchants (*15*) imported considerable numbers of horses into western North America.

Whereas horses would generally be categorized as domestic commodities, Indigenous peoples often maintain different relationships with them. Lakota peoples attribute to horses a nationhood status equal to their own. The Lakota–horse relationship is thus one of great reverence, deeply embedded in their identity, spirituality, science, and cosmogony. Lakota peoples do not have concepts for "wild" and "domesticated." In fact, *Šungwakaŋ*—"the Horse Nation"—was neither controlled behind fences nor forced into breeding. Rather, the Lakota peoples strove to cultivate their environment and adapt their lifeways to ensure that *Šungwakaŋ* could live aligned with its natural systems. Within this nation-to-nation alliance, the horse enhanced the abilities of the Lakota with regard to hunting, mobility, healing, and more (*16*). Therefore, for the Lakota peoples, saying "our horse" never reflects ownership but rather responsibility for a sacred relative.

European colonization entirely altered Indigenous social dynamics, hierarchy, and lifeways, introducing profound changes to subsistence modes, movement, and warfare (*17*). Many Indigenous peoples within the Great Plains and American Southwest developed horse-based pastoral or hunting economies and expanded transcontinental networks of raiding and exchange. Some became militarily dominant polities that maintained autonomy and sovereignty into the end of the 19th century CE, with many maintaining this sovereignty today (*18*, *19*).

Historical models for the post-Columbian North American dispersal of horses and their integration into Indigenous cultures are almost exclusively derived from textual sources written by European observers dating largely to the 18th and 19th centuries CE [e.g., (*20*, *21*)]. These sources depict horses first spreading in appreciable numbers north from what is today the American Southwest after the Pueblo Revolt of 1680 CE, when Spanish settlers were temporarily expelled from much of New Mexico (*22*). Given that most of the continent north of New Mexico was terra incognita to European chroniclers, natural and cultural landscapes remained largely uncharacterized until the early 19th century CE (*23*). Furthermore, these Euro-American historic records are often rife with inaccuracies and strong anti-Indigenous biases, depreciating the fundamental relationship between Indigenous peoples and horses (*24*).

Despite representing a major source for understanding the timing and ways in which horses were managed, ridden, and integrated into early societies, archaeological remains of domestic horses from Indigenous contexts are also overlooked (*24*). In this study, we extensively surveyed existing archaeological collections to identify early historic horse specimens with potential for reconstructing early human–horse relationships across the American Southwest and Great Plains (Fig. 1). Together, DNA, archaeozoological, and stable isotope data support the introduction of

Spanish-sourced domestic horses into Indigenous societies across the plains before the first half of the 17th century CE.

Results

Indigenous societies incorporated horses before the Pueblo Revolt

Of 33 early American equid specimens, we successfully radiocarbon dated 29 and characterized a total of 27 genetically, along with six new specimens from Eurasia (producing nine ancient genomes with an average depth-of-coverage of 2.06× to 12.24×, with substantial genome-wide sequence data for seven additional horse specimens, 0.06× to 0.96×, plus one donkey genome, 1.32×) (Fig. 1). Zonkey software analyses (*25*) confirmed all specimens as horses, except NW36 from Chupaderos, Mexico, which is a donkey jennet (table S1). Although a plateau in the radiocarbon calibration curve prevents easy discrimination between horses dating between 1670 CE and the early 20th century CE, we identified three horses from North American Indigenous contexts conclusively predating the Pueblo Revolt. Near-infrared (NIR) spectrum analysis failed to detect any external contaminants that could have affected radiocarbon dating (materials and methods section 3). The three specimens include a juvenile horse burial from the site of Blacks Fork in southwestern Wyoming, an adult horse cranium from Kaw River, Kansas, and isolated skeletal elements from the site of Paa'ko, New Mexico, along with new analysis of a previously dated specimen from American Falls Reservoir, Idaho, dated to between 1597 and 1657 CE (*26*), which we also assessed with NIR spectroscopy (materials and methods section 3). Assuming that the historic reintegration of horses was bounded temporally by the first presence of European horses on the North American mainland (1519 CE), Bayesian radiocarbon modeling suggests a date of between 1516 and 1599 CE (2σ modeled range) for the initial adoption of horses by Indigenous societies in western North America, with a median boundary date of ~1544 CE (Fig. 1D and materials and methods section 2). Various models provided good measures of agreement (A_{model} and $A_{overall}$ > 80 in all cases), and excluding anomalous values did not meaningfully affect date estimates (materials and methods section 2).

Historic North American horses descend primarily from Spanish genetic sources

Molecular phylogeny revealed that historic and modern North American male horses carried Y-chromosomal haplotypes belonging to the "Crown group" (Fig. 2A), which became dominant within the past ~1500 years, following the increasing popularity of oriental stallions at the origin of most non-Asian domestic bloodlines today, including Arabians, Barbs, and Thoroughbreds (*27*). Mitochondrial phylogenetic inference also rejected maternal continuity from Late Pleistocene horses excavated both north and south of the North American ice sheets (Fig. 2B). Furthermore, BIONJ phylogenetic reconstruction based on autosomal variation at ~7.5 million nucleotide transversions supported a deep divergence between Late Pleistocene North American horses and all present and past lineages identified in Eurasia. This analysis placed both historic and modern North American horses within the genomic variation of modern domestic horses (Fig. 2C). Combined, these phylogenetic reconstructions portray historic and modern North American horses as mainly descending from domestic bloodlines that started spreading outside their native area of the Don-Volga region no earlier than 4200 years ago (*28*).

Admixture graph modeling did not show evidence of gene flow from Late Pleistocene into historic or modern North American horses (fig. S6.2). The individual ancestry profiles of North American horses were consistent with those found in recent domestic Eurasian bloodlines, sporadically including a minor possible contribution from Late Pleistocene North American horses or related lineages (<0.73%) (Fig. 2C). This ancestry was, however, not exclusive to historic or modern North American horses but instead shared across most Eurasian lineages, including a ~4000-year-old horse from Iberia, a ~5100-year-old horse from western Beringia, and several ancient domestic specimens such as a 1447 to 1621 CE sample from Iran (Belgheis). Therefore, the minor ancestry component detected likely reflects multiple ancient contacts between Eurasia and North America through the Beringian land bridge during the past 830,000 years, in line with previously reported studies (*26*) and also apparent in mitochondrial phylogenies (Fig. 2B).

To further characterize the main genetic sources of North American horses, we implemented the qpAdm modeling rotation scheme (*29*), considering either single or two-donor sources among 37 populations. These included

[1]Department of Anthropology, University of Colorado Boulder, Boulder, CO 80309, USA. [2]Museum of Natural History, University of Colorado Boulder, Boulder, CO 80309, USA. [3]Centre for Anthropobiology and Genomics of Toulouse (CAGT, CNRS UMR5288), University Paul Sabatier, Faculté de Médecine Purpan, 31000 Toulouse, France. [4]Oglala Lakota, Pine Ridge Reservation, SD 57770, USA. [5]Pawnee Nation of Oklahoma, Pawnee, OK 74058, USA. [6]Tribal Historian, Comanche Nation, Galindo Environmental Consulting LLC, Austin, TX 78757, USA. [7]Lakota, Pine Ridge Reservation, SD 57770, USA. [8]International Indian Treaty Council, San Francisco, CA 94103, USA. [9]Sicangu Lakota, Rosebud Indian Reservation, SD 57570, USA. [10]He'Sapa Unity Alliance Council of Elders, SD 57770, USA. [11]Cheyenne River Sioux Tribe (Lakota), Eagle Butte, SD 57625, USA. [12]Pueblo of Acoma, Acoma, NM 87034, USA. [13]Zoological Institute, Department of Environmental Sciences, University of Basel, 4051 Basel, Switzerland. [14]Department of Biotechnology, Abdul Wali Khan University, Mardan 23200, Pakistan. [15]Institute of Culture and Environment, Alaska Pacific University, Anchorage, AK 99508, USA. [16]Deg Xit'an (Athabascan), Shageluk Tribe of Interior Alaska, Shageluk, AK 99665, USA. [17]Kentucky Archaeological Survey, Western Kentucky University, Bowling Green, KY 42101, USA. [18]W.S. Webb Museum of Anthropology, University of Kentucky, Bowling Green, KY 42101, USA. [19]Conservation and Evolutionary Genetics Group, Estación Biológica de Doñana (EBD-CSIC), 41092 Sevilla, Spain. [20]Laboratorio de Paleontología y Paleobiología, Instituto Andaluz del Patrimonio Histórico, 41092 Sevilla, Spain. [21]Université Paris-Saclay, INRAE, AgroParisTech, GABI UMR1313, Jouy-en-Josas, 78350 Paris, France. [22]Musée de l'Armée, Hôtel des Invalides, 75007 Paris, France. [23]The Royal Danish Academy, Institute of Conservation, 1435 Copenhagen K, Denmark. [24]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064, USA. [25]University of Southampton Faculty of Arts and Humanities (Archaeology), Southampton SO17 1BF, UK. [26]Institute for Anthropological Research, 10000 Zagreb, Croatia. [27]Centre National de Recherche Scientifique, Muséum national d'Histoire naturelle, Archéozoologie, Archéobotanique (AASPE), CP 56, 75005 Paris, France. [28]Faculty of History, University of Oxford, Oxford OX1 2RL, UK. [29]Oxford Nizami Ganjavi Centre, Faculty of Oriental Studies, University of Oxford, Oxford OX1 2LE, UK. [30]Genoscope, Institut de Biologie François Jacob, CEA, CNRS, Université d'Evry, Université Paris-Saclay, 91000 Évry, France. [31]Biology Department, College of Science, Taif University, Taif 21944, Saudi Arabia. [32]Zoology Department, College of Science, King Saud University, Riyadh 12372, Saudi Arabia. [33]Biology Department, College of Science, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia. [34]Department of Information Systems, College of Applied Sciences, Almaarefa University, Riyadh 13713, Saudi Arabia. [35]Department of Anatomy, Histology and Embryology, Faculty of Veterinary Medicine, University of Zagreb, 10000 Zagreb, Croatia. [36]Department of Archaeology, Faculty of Humanities and Social Sciences, University of Zagreb, 10000 Zagreb, Croatia. [37]School of History, Classics and Archaeology, University of Edinburgh, Edinburgh EH8 9AG, UK. [38]isoTROPIC Research Group, Max Planck Institute for Geoanthropology, 07745 Jena, Germany. [39]Department of Anthropology, University of Utah, Salt Lake City, UT 84112, USA. [40]Archaeological consultant, Omaha, NE 68131, USA. [41]Department of Anthropology, University of Wyoming, Laramie, WY 82071, USA. [42]Department of Anthropology, University of New Mexico, Albuquerque, NM 87131, USA. [43]Department of Anthropology, Texas A&M University, College Station, TX 77840, USA. [44]SWCA Environmental Consultants, Inc., Sheridan, WY 82801, USA. [45]Department of Entomology, University of Arizona, Tucson, AZ 85721, USA. [46]Department of Geography, University of Colorado Boulder, Boulder, CO 80309, USA. [47]Department of Geological Sciences, University of Cape Town, Rondebosch 7700, South Africa. [48]Department of History, Anthropology, Philosophy, Political Science, and Department of Spanish, Adams State University, Alamosa, CO 81101, USA. [49]Universidad Nacional de la Patagonia Austral, Unidad Académica Río Gallegos (ICASUR), Laboratorio de Arqueología Dr. Luis A. Borrero, CONICET, 9400 Río Gallegos, Santa Cruz, Argentina. [50]Quaternary Palaeontology Program, Royal Alberta Museum, Edmonton, AB T5J 0G2, Canada. [51]Department of Ecology and Evolutionary Biology and Howard Hughes Medical Institute, University of California, Santa Cruz, CA 95060, USA. [52]Department of Earth System Science, University of California, Irvine, CA 92697, USA. [53]Dartmoor Hill Pony Association, Corndonford Farm, Poundsgate, Devon TQ13 7PP, UK. [54]Department of Archaeology, University of Exeter, Exeter EX4 4QE, UK. [55]Department of Agriculture and Industry, Sul Ross State University, Alpine, TX 79832, USA. [56]Department of Virology, Florida Department of Health, Jacksonville, FL 32202, USA. [57]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden. [58]Xeni Gwet'in First Nations Government, 150-Milehouse, BC V0K 2G0, Canada. [59]McCrory Wildlife Services Ltd., New Denver, BC V0G 1S1, Canada. [60]Center for Animal Breeding and Genetics, Department of Biosystems, KU Leuven, 3001 Leuven, Belgium. [61]Department of Animal Science, UF Genetics Institute, University of Florida, Gainesville, FL 32610, USA. [62]Plateforme GeT-PlaGe, Génome et Transcriptome, US1426, Centre INRAe Occitanie, 31326 Auzeville, France. [63]UA Accelerator Mass Spectrometry Laboratory, University of Arizona, Tucson, AZ 85721, USA. [64]Department of Anthropology, University of Oklahoma, Norman, OK 73019, USA. [65]Oklahoma Archeological Survey, University of Oklahoma, Norman, OK 73019, USA. [66]Department of Archaeology, Max Planck Institute for Geoanthropology, 07745 Jena, Germany.
***Corresponding author. Email: william.taylor@colorado.edu (W.T.T.T.); ludovic.orlando@univ-tlse3.fr (L.O.)** †These authors contributed equally to this work. ‡Deceased.

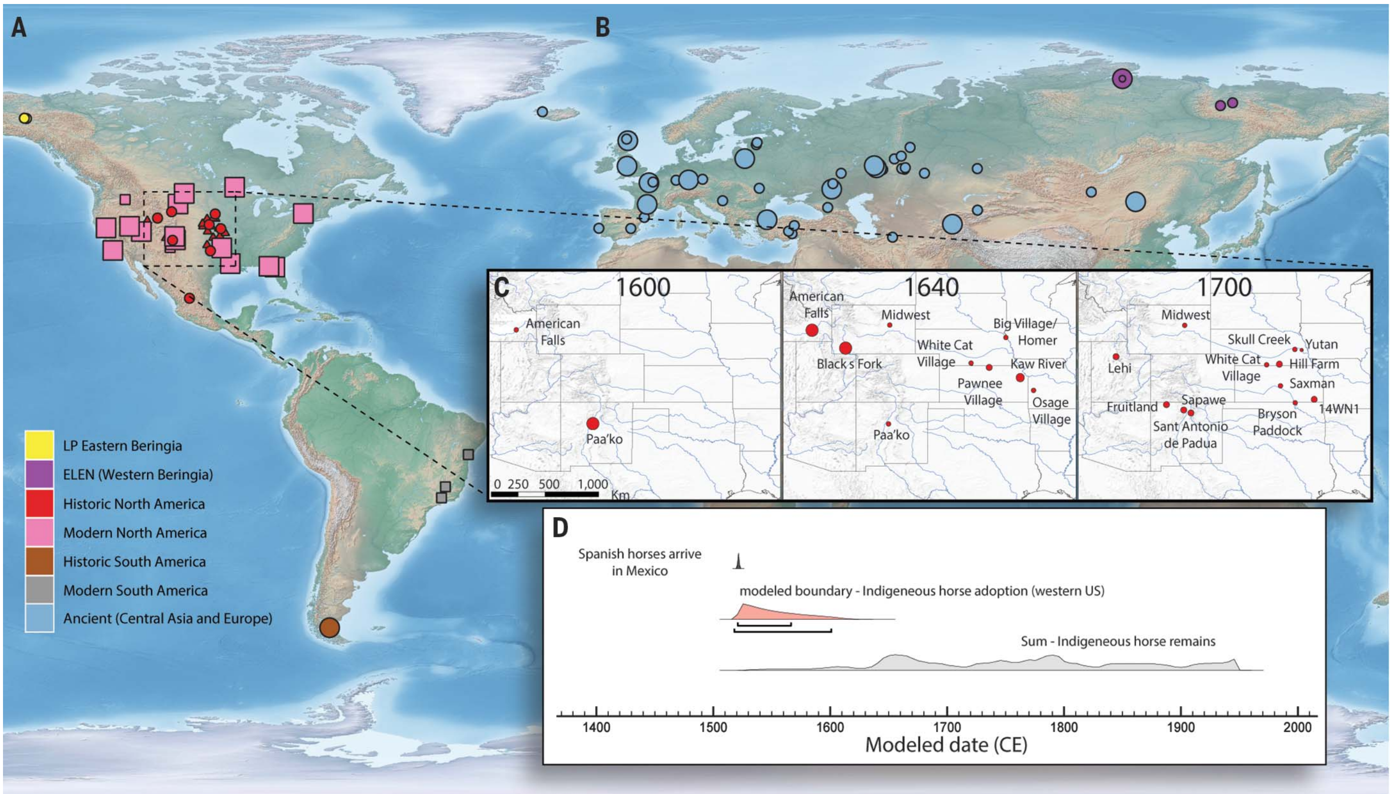


Fig. 1. Sample spatial and chronological distribution. (**A**) American samples. Processed archaeological and/or museum specimens are shown as red triangles, with those successfully sequenced shown as a plain circle; modern specimens are indicated with squares. LP, Late Pleistocene [data from Vershinina *et al.* (*26*)]; ELEN, *Equus lenensis* (Siberia). (**B**) Ancient Eurasian samples in the comparative panel (table S1). Horses from western Beringia are colored in purple, the remaining in blue. Modern genomes are not projected on the map but are provided in table S1. (**C**) Geographic visualization of Bayesian radiocarbon dates for early horse dispersals. Produced in OxCal assuming a uniform prior, with results transferred and visualized in ArcGIS. For each specimen, point diameter corresponds to the portion of total probability distribution within each time slice. Note that presence or absence of a dot does not necessarily indicate occupation during a given time slice (see materials and methods section 2 for detailed chronological modeling of sites thought to predate 1680 CE). (**D**) Modeled radiocarbon boundary for introduction of Spanish-sourced domestic horses into Indigenous societies in the western United States, based on analyzed radiocarbon dates.

Late Pleistocene North American horses as well as a representative panel of both modern and ancient domestic horses from around the globe. This analysis rejected Late Pleistocene North American horses as a possible source for both historic and modern North American horses but supported domestic bloodlines with almost exclusively European origins. Moreover, historic North American horses showed strong genetic affinities to three ancient domestic horses from Spain (Jal5885), Iran (Belgheis), and France (Inva22), dated to between the 15th and 18th centuries CE. This was true for all specimens except the one from Fort Boonesborough (Boones1/Kentucky, 1778 CE), which showed greater genetic affinity to British horses from the 18th century CE (Witter Place) (Fig. 3A). Notably, the influence of British bloodlines is also apparent among modern North American horses, which are commonly modeled as mixtures between Spanish-like (Galiceño and Cartujano) and British-like sources (Thoroughbreds and Welsh, Shetland, and Dartmoor Hill ponies) (table S3 and Fig. 3B). This indicates a temporal shift in the genetic composition of North American horses tracing new inputs from growing British and, later, American presence in eastern North America, and the integration of these horses into Indigenous spheres of interaction. Historic North American horses show greater relatedness to the historic Iberian specimen (Jal5885) than to modern Welsh bloodlines, but modern North American horses show increased Welsh relatedness (Fig. 3, B, C, and E). This is true for all except for feral horses isolated in the Santa Cruz Islands (STCZ3322 and STCZ3327), consistent with their alleged Spanish historical origins. Finally, both historic and most modern North American groups remain closer genetically to Jal5885 than to Icelandic horses (Fig. 3D), and qpAdm modeling revealed ancestry profiles incompatible with genetic contribution from a Viking specimen dating to between the 9th and 11th centuries CE (VHR102) (supplementary materials section 6).

The genomic analyses described above focus on the minute fraction of the genome that is different across all horse lineages investigated. However, the Lakota find key instruction in the >99% of the genome fraction that appears common to all.

Pre–Pueblo Revolt contribution of horses to Indigenous beliefs, trade, and transport networks

Archaeological specimens dated to the early 17th century CE show pathological and osteological evidence of care, management, and use in transport. A horse phalanx from an Indigenous-affiliated context at Paa'ko Pueblo, New Mexico, and a metacarpal from American Falls Reservoir, Idaho, demonstrate the presence of horses among Native communities as far north as Idaho by the first half of the 17th century CE. Furthermore, a foal from Blacks Fork, Wyoming, exhibits entheseal ossification of the nuchal ligament attachment at the rear of the skull at levels typically and exclusively found in horses used for transport or kept in confinement (*30*, *31*). It also features a severe, healed cranial fracture that may have resulted from a kick caused by confinement in close

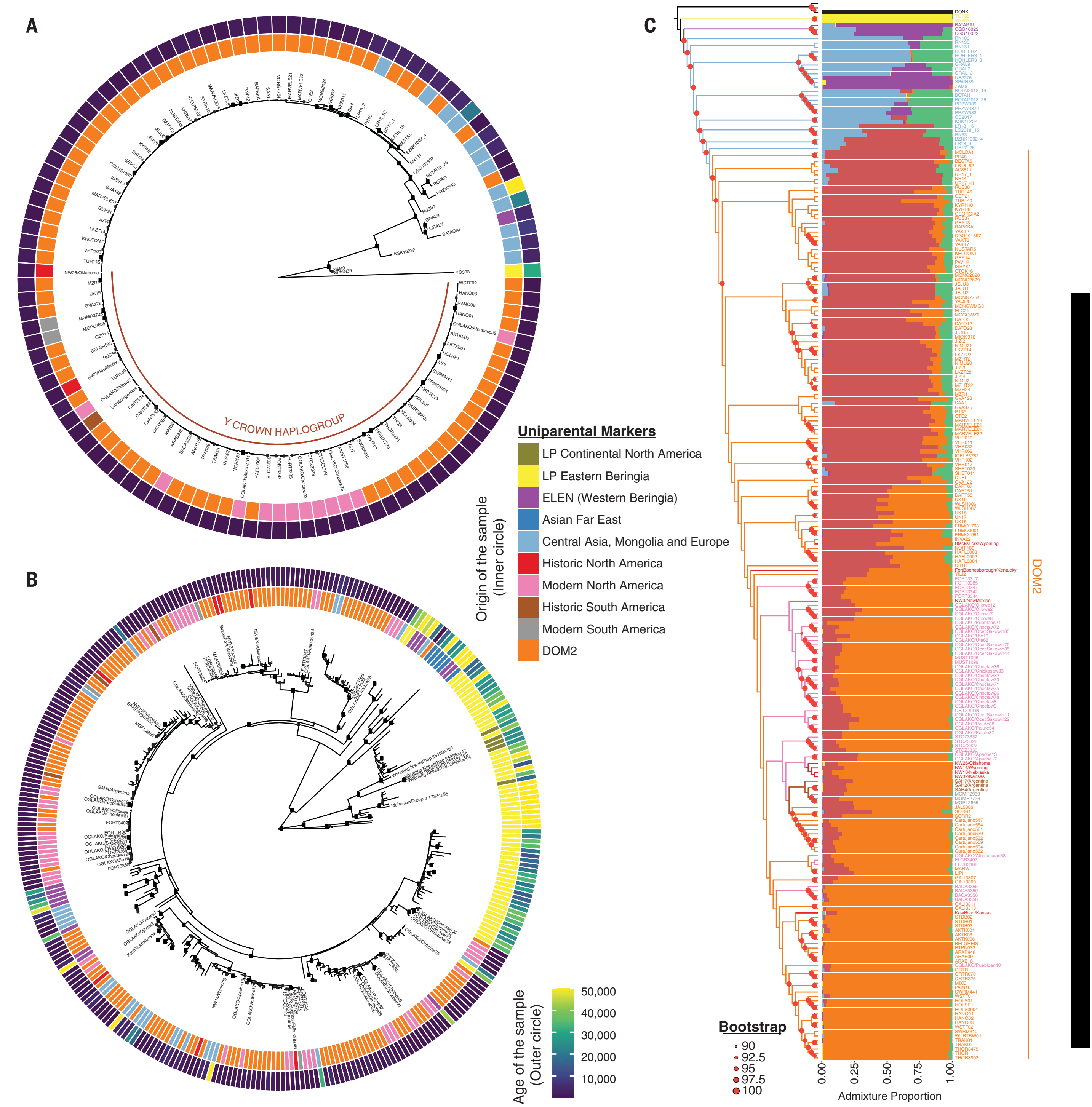


Fig. 2. Phylogenetic affinities. (**A**) Y-chromosomal maximum-likelihood tree (*N* = 110; 5244 nucleotide transversions). (**B**) Mitochondrial maximum-likelihood tree (*N* = 340; 16,406 base pairs). (**C**) Neighbor-joining tree based on ~7.5 million nucleotide transversions (*N* = 241; left), and the corresponding genetic ancestry profiles (right). The ancestry proportions for *K* = 7 genetic components were estimated using Struct-f4 (*28*). Sample names written as "OGLAKO/Apache" represent modern North American horses culturally associated with Apache communities. LP, Late Pleistocene; ELEN, *Equus lenensis* (Siberia); DOM2, modern domestic lineage descending from a lower Don-Volga genetic source expanding to Eurasia after 4200 years ago.

proximity to other horses (materials and methods section 4). Moreover, this foal was recovered in ritual features alongside coyotes (*32*), indicating that horses were already integrated into existing social and ceremonial traditions by the first half of the 17th century CE. Nuchal ossification along with characteristic damage linked to the use of a metal bit, including erosion of the anterior cementum and enamel of the lower second premolar, were also found in the horse from Kaw River, Kansas, dating to the same period (Fig. 4) (*33*). Several other pathological features reflect the use of a metal curb bit of the type used by early Spanish colonists and later Mexican and

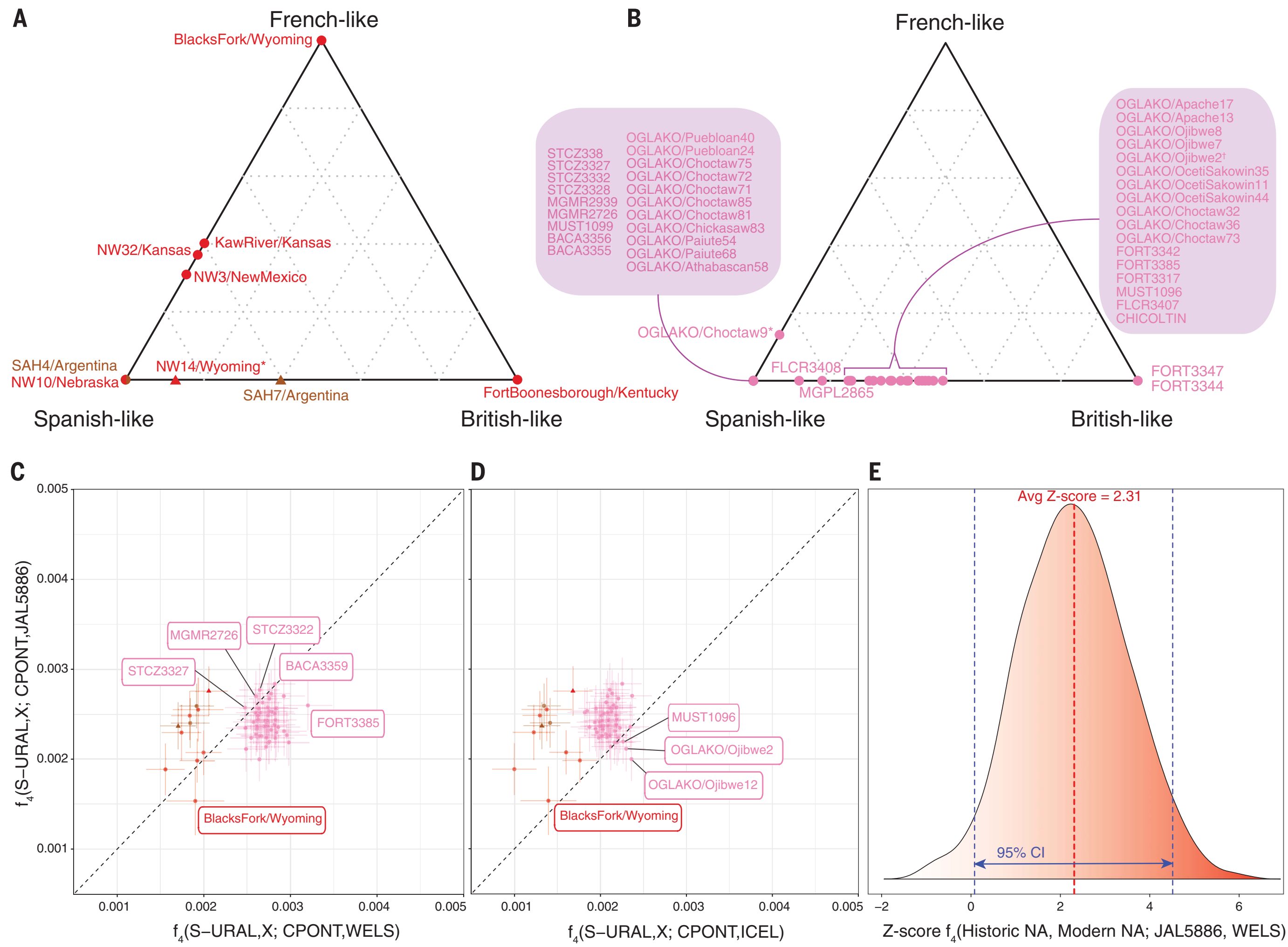


Fig. 3. Temporal changes in the genomic makeup of American horses. (**A**) Ternary plots showing ancestry combinations of the historic horse genomes from North America. (**B**) Ternary plots showing ancestry combinations of the modern horse genomes from North America. Ancestry proportions were estimated using qpAdm modeling and rotating 37 possible donor sources. Models with highest *P* values are shown where multiple models could not be rejected. Asterisks depict two samples for which the ancestry profiles presented correspond to the second best qpAdm model (table S3), while the cross accompanying OGLAKO/Ojibwe2 indicates that its best qpAdm model involves Galiceño and Icelandic horses as donor populations, instead of British sources (supplementary materials section 6). (**C**) Relative genetic affinities of historic and modern North American horses to Jal5885 versus Welsh horses. Genetic affinities are estimated using f_4-statistics in the form of (S-URAL, X; C-PONT, Jal5885 or Welsh), where X represents historic and/or modern horses from North America, and C-PONT horses from the third millennium CE showing the closest genomic affinities to modern DOM2 domesticates. Triangles show the two historic samples that returned modern radiocarbon dates (NW14/Wyoming and SAH7/Argentina). (**D**) Relative genetic affinities of historic and modern North American horses to Jal5885 versus Icelandic horses. The calculations are the same as in (C), except that modern Welsh horse genomes were replaced by genomes from modern Icelandic horses and the VHR102 Viking sample (850 CE to 1050 CE). An additional qpAdm analysis considering modern Icelandic horses and the sample VHR102 as separate donors of genetic ancestry confirmed that the specimen OGLAKO/Ojibwe2 received introgression from modern Icelandic horses, ruling out genetic contribution from relict populations of Viking horses (supplementary materials section 6). Error bars correspond to twice the standard error estimated from jackknifing. (**E**) Distribution of Z-scores for f_4-statistics of the form (Historic North American Horse, Modern North American Horse; Jal5885, Welsh).

Navajo craftsmen, such as the fracturing of the upper palate caused by the curb (*34*) and arthritis of the temporomandibular joint. Combined, these analyses indicate that early plains horses were already used for mounted riding.

Strontium isotope values for Blacks Fork, reflecting prenatal (dP4, $^{87}Sr/^{86}Sr = 0.70970$) and postnatal (M1, $^{87}Sr/^{86}Sr = 0.70969$) values for the foal, are directly in line with published values for modern fauna from the Green River Basin, where the specimen was found (0.70950 to 0.71000) (*35*, *36*). Although similar values characterize some stretches of the Colorado River tributary system farther to the south (*37*), the Blacks Fork values rule out most of Wyoming, including the border with Utah (*38*). Additionally, strontium isotope results from the Kaw horse, Kansas, are consistent with available values from northeastern Kansas (*39*) but indicate some mobility during the recorded sequence ($^{87}Sr/^{86}Sr = 0.70905$ near the crown, versus $^{87}Sr/^{86}Sr = 0.70930$ near the root), which spans a 12- to 14-month period across the horse's fourth and fifth year of life, according to mineralization schedules (*40*). When considered alongside stable isotopes of oxygen ($\delta^{18}O$) and carbon ($\delta^{13}C$), strontium isotope values suggest that the Kaw horse spent part of its life farther north, indicating gradual movement from an area matching reference data from South Dakota, Nebraska, and Iowa (Fig. 4 and materials and methods section 5). Therefore, isotope values from both early horses with complete dentition suggest that the animals were either raised and managed

 https://avxhm.se/blogs/hill0

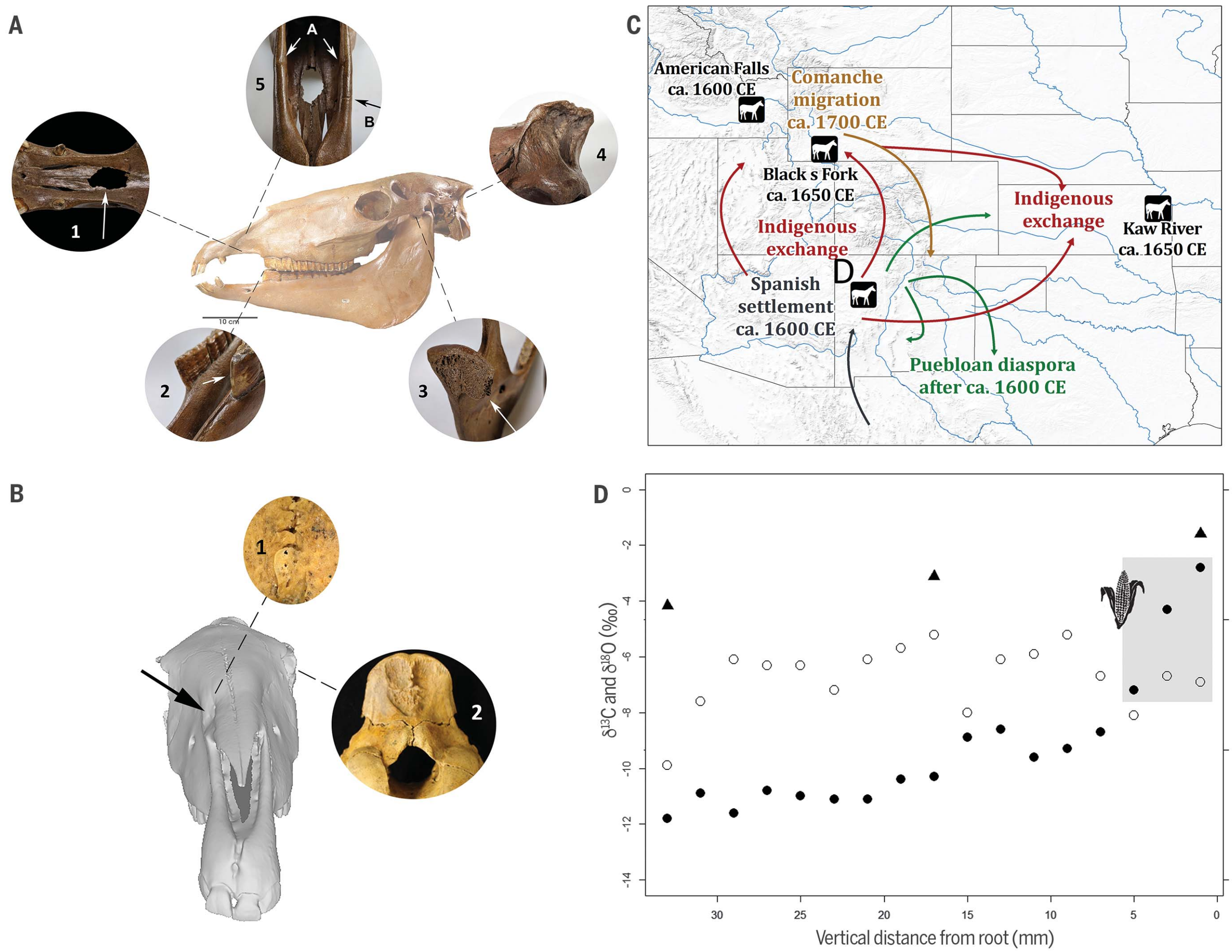


Fig. 4. Horse herding in the 17th century CE. (**A**) Osteological indicators of bridling and riding: (1) palatal damage from curb bit; (2) bit damage to anterior margin of lower second premolar; (3) osteoarthritis at the temporomandibular joint; (4) entheseal changes to the nuchal ligament attachment site; and (5) remodeling of the premaxilla. (**B**) Osteological indicators of human activity: (1) Healed kick fracture in young foal, suggesting health care. The inset shows healing at 50× magnification. (2) Ossification of the nuchal ligament, indicating use in transport or confinement and/or tethering. (**C**) Modeled northward dispersal of horses. Model based on archaeological discoveries and consideration of historically documented cultural developments and migrations. (**D**) Isotopic evidence for foddering. Stable carbon and oxygen and strontium isotope sampling locations for the Kaw River horse, as measured in millimeters from the root (lower-right third molar). Sampling locations are represented on the *x* axis as the midpoint of a 2-mm section (e.g., the section from 0 to 2 mm is shown as "1" on the graph). Filled triangles show strontium isotope values, with stable carbon values represented by filled circles, and stable oxygen values as open circles. [Kaw horse image credit: E. Scott]

locally (Blacks Fork) or within a territory extending even farther away from the European colonial sphere (Kaw).

A $\delta^{13}C$ spike toward the end of the sampled section of the Kaw molar shows a strong input of C_4-pathway plants, potentially reflecting winter foddering with the Indigenous domestic crop maize (materials and methods section 5), a traditional practice among Plains groups, including the Pawnee. Other situations, such as herd movement south and west to rangeland higher in C_4 grasses or during events such as seasonal bison hunts, could account for the enriched value but are not consistent with strontium and oxygen isotope data. Pawnee (*Chaticks si Chaticks*) villages typically relied on stores of maize to subsist through the winter (*41*), and the practice of winter foddering horses with maize in the northern Missouri River region is documented ethnographically during the 18th and 19th centuries CE, including among the Pawnee (*42*, *43*). Regardless of whether the Kaw specimen was affiliated with ancestral Pawnee or another Central Plains nation, our results indicate the presence of horses in Indigenous cultural and economic systems of the Missouri River drainage during the first half of the 17th century CE.

Discussion

Our archaeological analyses show the dispersal of domestic horses from Spanish settlements in the American Southwest to the northern Rockies and central Great Plains by the first half of the 17th century CE at the latest. They provide evidence of local raising and veterinary care of horses, likely foddering with domestic maize, and use of horses in transport by Indigenous peoples by this time. A directly dated radiocarbon specimen from Paa'ko Pueblo in northern New Mexico shows that horses reached the region via Indigenous groups before Spanish colonization of the American Southwest, as previously hypothesized (*44*, *45*). Moreover, our new temporal framework shows that horses were present across the plains long before any documented European presence in the Rockies or the central plains. Despite their

Iberian genetic makeup and earlier arguments attributing one of these horses to Spanish exploration (*46*), strontium, carbon, and oxygen isotope results suggest that these animals were raised and died locally. Osteological analyses also provide some indication that horse dispersal was tied to broader economic links; the Kaw River horse was indeed controlled with a European-style metal curb bit, and the Blacks Fork horse body was cut with metal tools (*32*). Therefore, a possible mechanism of horse dispersal was exchange across Indigenous networks at the margins of New Mexico and the American Southwest (*47*). Once acquired, horses could have moved via ancient trade routes based on kinship ties and social networks established throughout the Great Plains and Rockies millennia before European contact (*48*). The dispersal of Puebloan groups toward the northeast, from Spanish New Mexico into western Kansas, provides another mechanism for the transmission of domestic horses into the Central Plains.

Our findings have deep ramifications for our understanding of social dynamics in the Great Plains during a period of disruptive social changes for Indigenous peoples. The area of southwestern Wyoming including Blacks Fork is considered to be a homeland for the Shoshonean-speaking ancestral Comanche (Nʉmʉnʉʉ), who migrated to the southern plains of Texas, New Mexico, Colorado, and Oklahoma before the early 18th century CE (*49*, *50*). The drive to acquire horses from Spanish New Mexico is often cited as a likely mechanism for this transcontinental movement (*50*). However, our new data—which align with some Comanche and Shoshone oral accounts (*51*) (materials and methods section 7)—suggest that ancestral Comanche had already integrated horse raising, ritual practices, and transport into their lifeways at least a full half century before their southward migration, effectively moving to the southern plains as horse herders. Once in the southern plains, the Comanche were able to marshal these advantages, along with their herding and equestrian skills, to build an empire on horse and bison trade by the middle of the 18th century CE (*19*). By the time Europeans arrived in southwestern Wyoming, the area was already a critical "secondary diffusion center" for horse transmission to Northern Plains groups (*17*, *52*). Considering the small body of archaeological data available, our findings raise the possibility of rapid, non-European transmission of horses farther northward, including the Columbia Plateau, the Canadian Rockies, and the middle and upper Missouri regions.

DNA data from horses currently caretaken and historically protected by the Oglala Lakota were included in this study. All protocols for the transmission of sacred and traditional knowledge were followed, as outlined by our Lakota Elder Knowledge Keeper Internal Review Board. They provide further context of the findings, which clarifies the Lakota culture and their relationship with the horse. Chief Joe American Horse states: "Horses have been part of us since long before other cultures came to our lands, and we are a part of them. The Horse Nation is our relative. We always protect our relatives and the next seven generations. We stand with the horse and we will always do so however it has evolved through its journey. That is what being Lakota is" (original quote, in Lakota, provided in materials and methods section 7). This study established that Indigenous peoples were living and interacting with the horse before the Pueblo Revolt of 1680 CE, which was the earliest date accepted by Western science. However, current genetic evidence shows that the horses caretaken by Indigenous peoples from as early as the first half of the 17th century CE do not share an excess of genetic ancestry with Late Pleistocene North American horses. Given that the Horse Nation is foundational to Lakota lifeways (*16*), one possible implication of this finding is that relationships of the kind developed by Lakota peoples could have already been in place by the Late Pleistocene. Such life management practices may even have extended to other members of the horse family at that time. Testing these implications requires further paleontological, archaeological, genetic, and ethnographic research. Dr. Antonia Loretta Afraid of Bear-Cook adds: "The Horse Nation always chose their own mates. Bringing in new blood is a replenishment and renewal process of life that we celebrate. It strengthens our *wé* (the life force from our blood). No matter how our horses may have transformed, or where they are around the world, we will always call to them. Together we are home" (original quote, in Lakota, provided in materials and methods section 7). This study demonstrates that colonization did not just drastically affect Indigenous peoples but also their horses, whose genetics captures an ancestry shift from Spanish to British bloodlines. Dr. Antonia Loretta Afraid of Bear-Cook thus reminds us that the current genetic makeup of the horse does not change the Lakota responsibility toward *Šungwakaŋ*. In fact, for the Lakota and other Indigenous peoples with deep ancestral relationships with the horse, their evolution was not to be feared or controlled, but rather something to be respected. Protecting horses now, regardless of their origins, is in fact protecting the Lakota and other Indigenous lifeways. A commitment to protecting these is a commitment to protect all life.

REFERENCES AND NOTES

1. B. J. Macfadden, *Science* **307**, 1728–1730 (2005).
2. L. Orlando *et al*., *Nature* **499**, 74–78 (2013).
3. B. Kooyman, L. V. Hills, P. McNeil, S. Tolman, *Am. Antiq.* **71**, 101–121 (2006).
4. S. D. Webb, C. A. Hemmings, *BAR International Series* **1560**, 11 (2006).
5. J. B. Wheat, *West. Can. J. Anthropol.* **2**, 169–177 (1982).
6. Y. Running Horse Collin, The Relationship Between the Indigenous Peoples of the Americas and the Horse: Deconstructing a Eurocentric Myth, Zenodo (2017); https://doi.org/10.5281/zenodo.7217326.
7. T. J. Murchie *et al*., *Nat. Commun.* **12**, 7120 (2021).
8. J. Haile *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22352–22357 (2009).
9. I. B. Enghoff, *Hunting, Fishing and Animal Husbandry at the Farm Beneath the Sand, Western Greenland* (Museum Tusculanum Press, 2003).
10. M. Kuitems *et al*., *Nature* **601**, 388–391 (2022).
11. B. Wallace, *J. North Atl.* **2**, 114–125 (2009).
12. P. Mitchell, *Horse Nations: The Worldwide Impact of the Horse on Indigenous Societies Post-1492* (Oxford Univ. Press, 2015).
13. J. Bowen, S. T. Andrews, "The starving time at Jamestown: Faunal analysis of pit 1, pit 3, the bulwark ditch, ditch 6, ditch 7, and midden 1," Report submitted to the Association for the Preservation of Virginia Antiquities, Jamestown Rediscovery, Williamsburg, VA (2000).
14. R. L. Jones, *Can. Hist. Rev.* **28**, 125–155 (1947).
15. I. V. Ovchinnikov *et al*., *PLOS ONE* **13**, e0200795 (2018).
16. M. Hunska Tašunke Icu (J. American Horse) *et al*., "Standing for Unči Maka (Grandmother Earth) and All Life: An Introduction to Lakota Traditional Sciences, Principles and Protocols and the Birth of a New Era of Scientific Collaboration" (Tech. Ser. No. 42, Maxwell Museum of Anthropology, Univ. of New Mexico, 2023).
17. J. C. Ewers, *The Horse in Blackfoot Indian Culture, with Comparative Material from Other Western Tribes* (Smithsonian Institution Press, 1969).
18. P. Hämäläinen, *Lakota America: A New History of Indigenous Power* (Yale Univ. Press, 2019).
19. P. Hämäläinen, *The Comanche Empire* (Yale Univ. Press, 2008).
20. C. Wissler, *Am. Anthropol.* **16**, 1–25 (1914).
21. F. Haines, *Am. Anthropol.* **40**, 429–437 (1938).
22. S. L. Olsen, in *Wild Equids: Ecology, Management, and Conservation*, J. I. Ransom, P. Kaczensky, Eds. (Johns Hopkins Univ. Press, 2017), pp. 105–120.
23. J. L. Kessell, *Whither the Waters: Mapping the Great Basin from Bernardo de Miera to John C. Frémont* (Univ. of New Mexico Press, 2017).
24. W. T. T. Taylor *et al*., *Am. Antiq.* **86**, 465–485 (2021).
25. M. Schubert *et al*., *J. Archaeol. Sci.* **78**, 147–157 (2017).
26. A. O. Vershinina *et al*., *Mol. Ecol.* **30**, 6144–6161 (2021).
27. S. Felkel *et al*., *Sci. Rep.* **9**, 6095 (2019).
28. P. Librado *et al*., *Nature* **598**, 634–640 (2021).
29. É. Harney, N. Patterson, D. Reich, J. Wakeley, *Genetics* **217**, iyaa045 (2021).
30. W. T. T. Taylor, J. Bayarsaikhan, T. Tuvshinjargal, *Antiquity* **89**, 854–871 (2015).
31. R. Bendrey, *Vet. Zootech.* **41**, 25–31 (2008).
32. C. A. Thornhill, *Plains Anthropol.* **66**, 58–73 (2020).
33. R. Bendrey, *J. Archaeol. Sci.* **34**, 1036–1050 (2007).
34. V. Onar, H. Alpak, G. Pazvant, A. Armutak, A. Chroszcz, *Rev. Med. Vet. (Toulouse)* **163**, 139–146 (2012).
35. J. N. Fenner, C. D. Frost, *J. Geochem. Explor.* **102**, 149–156 (2009).
36. A. C. Doebbert *et al*., *Chem. Geol.* **380**, 172–189 (2014).
37. L. A. Chesson, L. O. Valenzuela, G. J. Bowen, T. E. Cerling, J. R. Ehleringer, *Rapid Commun. Mass Spectrom.* **25**, 3713–3722 (2011).
38. E. L. Gross, P. J. Patchett, T. A. Dallegge, J. E. Spencer, *J. Geol.* **109**, 449–461 (2001).
39. C. Widga, J. D. Walker, A. Boehm, *Open Quat.* **3**, 4 (2017).
40. K. A. Hoppe, S. M. Stover, J. R. Pascoe, R. Amundson, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **206**, 355–365 (2004).
41. J. O'Shea, *Cent. Plains Archaeol.* **1**, 49–107 (1989).
42. D. E. Worcester, *Pac. Hist. Rev.* **14**, 409–417 (1945).
43. G. Weltfish, *The Lost Universe: Pawnee Life and Culture* (Univ. of Nebraska Press, 1977).
44. J. D. Forbes, *Southwest. J. Anthropol.* **15**, 189–212 (1959).
45. N. Blackhawk, *Violence over the Land: Indians and Empires in the Early American West* (Harvard Univ. Press, 2009).
46. J. A. Lockwood, R. Kumar, D. G. Eckles, *Am. Entomol.* **40**, 210–216 (1994).

47. M. Cowdrey, N. Martin, J. Martin, N. Williamson, *Bridles of the Americas: Horses & Bridles of the American Indians* (Hawk Hill Press, 2012).
48. M. D. Mitchell, *Crafting History in the Northern Plains: A Political Economy of the Heart River Region, 1400–1750* (Univ. of Arizona Press, 2013).
49. M. D. Mitchell, *Antiquity* **78**, 115–126 (2004).
50. G. Betty, *Comanche Society: Before the Reservation* (Texas A&M Univ. Press, 2005).
51. Å. Hultkrantz, *Z. Ethnol.* **93**, 49–72 (1968).
52. T. Binnema, *Common and Contested Ground: A Human and Environmental History of the Northwestern Plains* (Univ. of Oklahoma Press, 2001).

ACKNOWLEDGMENTS

We are grateful to the scholars, elders, and leaders who make up the Lakota Review Team and who supervised the Oglala Institutional Review Board process for this research, as well as their counterparts from other Indigenous nations who supported this research and interpretation. *Wopila Tanka* to the Horse Nation and each of the Indigenous nations who served as cultural caretakers for these horses. Special thanks to M. Sims, S. Olsen, and C. Beard of the KU Natural History Museum for facilitating access to the Kaw specimen (KUVP 347; imagery provided courtesy of E. Scott); to B. Peecook and A. Commendador for facilitating access to the American Falls specimen (IMNH 71004/25895, a Bureau of Reclamation specimen under the care of the Idaho Museum of Natural History, Idaho State University); and to D. Walker and the University of Wyoming Archaeological Repository (Office of the Wyoming State Archaeologist, 1000 E. University Ave., Dept. 3431, Laramie, WY 82071, USA) for providing imagery and facilitating access to collections at Blacks Fork (48SW8319), located in Flaming Gorge National Recreation Area, managed by the USDA–Ashley National Forest. Analysis of Blacks Fork material was completed with permission from the USDA–Ashley National Forest, coordinated through J. Rust, Heritage Program Leader. Additional thanks to B. Britt (the Maxwell Museum of Anthropology), D. Gifford-Gonzalez, and K. Kjær for facilitating research and access to additional collections. **Funding:** Research was funded by an award from the National Science Foundation (Collaborative Research: Horses and Human Societies in the American West, Awards 1949305, 1949304, and 1949283) and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 681605). Additional funding was provided by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (grant agreement 890702-MethylRIDE); the French Government "Investissement d'Avenir" FRANCE GENOMIQUE (ANR-10-INBS-09); the Universidad Nacional de la Patagonia Austral (PIP 29/A476); the Researchers Supporting Projects at King Saud University, Riyadh, Saudi Arabia (NSRSP2022R1), Taif University, Taif, Saudi Arabia (NSTURSP2022), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia (NSPNURSP2022), and Almaarefa University, Riyadh 13713, Saudi Arabia (NSTUMA/1); and the Marie Skłodowska Curie programme (101027750-HOPE). P.R. and M.L. thank the Max Planck Society for funding for the isotope analyses. **Author contributions:** Designed and coordinated the study: W.T.T.T., E.L.J., and L.O. Performed radiocarbon dating: J.S. and G.H. Performed ancient DNA laboratory work: L.Chau., L.T.-C., S.S., A.S.-O., A.F., N.K., C.D.S., X.L., S.W., and Y.R.H.C., with input from L.O. Performed ancient DNA computational analyses: P.L. and L.O. Provided material, reagents, and methods: W.T.T.T., P.R., B.L.M., N.O., J.A.L., E.B.-S., E.B., L.Char., E.R., T.D., K.G., A.O.V., J.W., P.R.Š., M.M., I.S., J.-M.A., A.Perd., S.A., A.H.A., K.A.S.A.-R., T.T.V., M.B., E.S., C.N., J.R.B., C.A.T., W.E.H., C.G., J.B.B., C.I.B.-O., C.R., M.Sp., B.S., J.S., A.O., L.P.R., K.P., M.So., A.W., W.M., G.L., S.B., C.E., C.D., O.B., P.W., G.H., S.T., B.B., E.L.J., Y.R.H.C., and L.O. Interpreted data: W.T.T.T., P.L., J.A., C.W., J.B.-C., J.R.B., C.A.T., V.M., A.Perr., W.E.H., J.L.C., P.L.R., S.G.B., J.B.B., C.R., M.Sp., G.H., S.T., B.B., P.R., E.L.J., Y.R.H.C., and L.O. Wrote the supplementary text: W.T.T.T., P.L., M.M., E.S., V.M., A.Perr., C.N., S.G.B., J.B.B., C.I.B.-O., I.A.H., S.T., B.B., E.L.J., Y.R.H.C., and L.O., with input from all coauthors. Wrote the article: W.T.T.T., B.B., E.L.J., Y.R.H.C., and L.O., with input from all coauthors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The sequence data generated in this study are available for download at the European Nucleotide Archive (accession no. PRJEB56773). The repository information for each individual sample and all other data used in this study are included in table S1.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adc9691
Materials and Methods
Figs. S1.1 to S6.4
Tables S1.1 to S3.3, S5.1, and S6.1 to S6.3
References (*53*–*124*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 15 May 2022; accepted 14 February 2023
10.1126/science.adc9691

GALAXIES

A cosmic stream of atomic carbon gas connected to a massive radio galaxy at redshift 3.8

Bjorn H. C. Emonts[1]*, Matthew D. Lehnert[2,3], Ilsang Yoon[1], Nir Mandelker[4,5,6,7], Montserrat Villar-Martín[8], George K. Miley[9], Carlos De Breuck[10], Miguel A. Pérez-Torres[11,12,13], Nina A. Hatch[14], Pierre Guillard[3,15]

The growth of galaxies in the early Universe is driven by accretion of circum- and intergalactic gas. Simulations have predicted that steady streams of cold gas penetrate the dark matter halos of galaxies and provide the raw material necessary to sustain star formation. We report a filamentary stream of gas that extends for 100 kiloparsecs and connects to the massive radio galaxy 4C 41.17. We detected the stream using submillimeter observations of the $^3P_1$ to $^3P_0$ emission from the [C I] line of atomic carbon, a tracer of neutral atomic or molecular hydrogen gas. The galaxy contains a central gas reservoir that is fueling a vigorous starburst. Our results show that the raw material for star formation can be present in cosmic streams outside galaxies.

The growth and evolution of galaxies are thought to be linked to the accretion of gas from the circum- and intergalactic media. Simulations have shown that a large fraction of the gas that accretes into the dark matter halos that surround galaxies is not shock-heated to high temperatures when passing through the accretion shock, but instead penetrates down into the halo as collimated steady flows of cold gas referred to as accretion streams (*1*, *2*). Although this process could occur throughout the history of the Universe, the simulations predicted that it peaked between redshift 2 and 4 (roughly 10 to 12 billion years ago) (*3*). Simulations that include these cold accretion streams match observations of high star formation rates in high-redshift galaxies (*4*).

If the time scales for cooling of the accreting gas are shorter than the time scales for mixing of hot and cold gas, then the streams are predicted to be stable against hydrodynamic instabilities, so additional gas from the circumgalactic medium cools and reinforces the streams (*5*). Observations at optical wavelengths have detected accreting streams around distant massive galaxies by means of the redshifted ultraviolet Lyman-α (Lyα) emission line of recombining hydrogen (*6*–*9*). This Lyα emission has been interpreted as arising from hydrogen gas that is cooling as it flows toward the adjacent galaxy (*10*–*12*). Although these results support the numerical simulations, Lyα emission only traces gas at temperatures of $\gtrsim 10^4$ K. Simulations lack the resolution to track the fragmentation that occurs when the gas cools to lower temperatures (*13*). This limits predictions about accreting gas at temperatures of 10 to 100 K, at which point hydrogen is expected to be in the molecular phase. Because molecular gas is the raw material that fuels star formation, it could potentially provide a direct link between cosmic gas accretion and galaxy growth.

Observations of 4C 41.17

We used the Atacama Large Millimeter/submillimeter Array (ALMA) to search for cold gas in and around the massive galaxy 4C

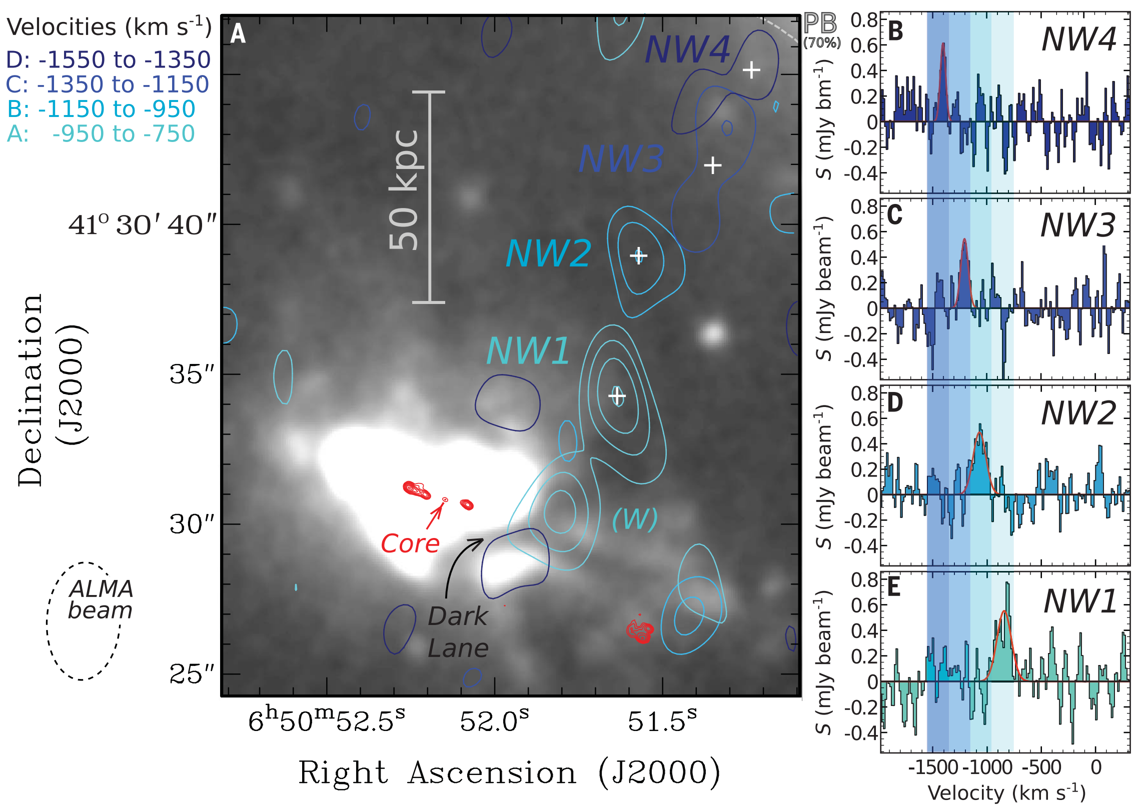


Fig. 1. [C I] emission around 4C 41.17. (**A**) Observed [C I] emission contours in velocity ranges indicated by the colors in the legend. The background grayscale image shows the Lyα halo (*17*). Contour levels are at 2σ, 3σ, 4σ, and 5σ, with σ = 0.011 Jy beam^{-1} km s^{-1} (where Jy is the jansky, 1 Jy = 10^{-26} W m^{-2} Hz^{-1}). No correction for primary beam (PB) response was applied, so fluxes are progressively attenuated with increasing distance from the center of 4C 41.17. The red contours show the radio source observed with MERLIN (*16*). Contour levels start at 0.33 mJy beam^{-1} and increase in steps of factor 2. The radio core at the center of 4C 41.17 and the dark lane in the Lyα image are labeled. Scale bar, 50 kiloparsecs at the redshift of 4C 41.17. The synthesized ALMA beam is shown by the dashed ellipse in the bottom left corner. (**B** to **E**) Spectra of [C I] at the locations northwest of 4C 41.17 marked by white crosses in (A) and labeled NW1, NW2, NW3, and NW4. These spectra were smoothed with a Hanning filter for display. Flux densities *S* have been corrected for the PB response. The red lines are Gaussian models fitted to each spectrum, and the shaded regions indicate the velocity ranges used in (A). Colors are the same as in the legend. Figure S4 shows [C I] emission at additional velocities, including in the central radio galaxy.

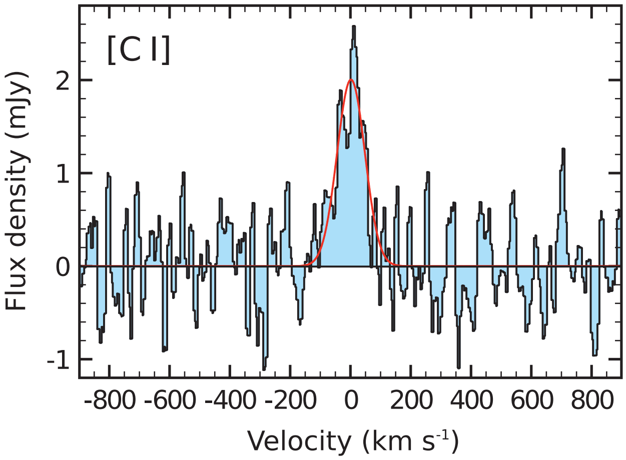


Fig. 2. Stacked spectrum of the [C I] emission in the stream. The spectrum includes only [C I]-emitting gas in the stream outside the Lyα halo, regions NW1, NW2, NW3, and NW4 in Fig. 1. The region labeled W in Fig. 1 (located west of the radio galaxy) was not included in this stack. The independent spectra of the regions NW1 to NW4 were coadded after shifting the spectra by aligning the velocities of the Gaussian models fitted to the [C I] line in each region (Fig. 1, B to E). The stacked spectrum uses channels of 14.5 km s^{-1}, and no Hanning smoothing was applied. The red line shows a Gaussian model fitted to the stacked spectrum, which was used to estimate the physical properties of the stream (table S1).

41.17 (also known as B3 0647+415) (*14*, *15*), which is at redshift z = 3.792 (*16*). 4C 41.17 consists of a group of small galaxies that are thought to be in the process of merging with a central radio galaxy to form a single massive galaxy (*14*, *15*, *17*). The central radio galaxy contains a supermassive black hole that generates jets visible at radio wavelengths (*14*). The entire 4C 41.17 system is surrounded by a giant halo of Lyα-emitting gas, which has a diameter of roughly 150 kpc (*17*). We used ALMA's most compact telescope configuration to increase its surface brightness sensitivity. The observing frequency was selected to search for the fine-structure $^3P_1$ to $^3P_0$ emission line of atomic carbon (hereafter [C I]), which has a rest-frame wavelength of 609 μm (492 GHz). The [C I] transition has an upper energy level (above the ground state) equivalent to 23.6 K and a critical density of $n_{\rm crit}$ ~ 500 cm^{-3}. It therefore traces molecular H_2 gas that has a temperature of 10 to 100 K and a volume density of $n_{\rm H2}$ less than several hundred particles per cubic centimetre (*18*). [C I] can be used as a tracer of molecular H_2 gas under a wide range of conditions, including in environments with a high cosmic ray flux (*18*, *19*) or where star formation has not yet commenced (*20*), although mass estimates rely on assumptions about the gas properties (supplementary text).

The observed distribution of [C I] emission in 4C 41.17 is shown in fig S4. By selecting emissions at different velocities, we imaged a string of [C I] blobs that align roughly in the north-south direction (Fig. 1). Their kinematics gradually transition from higher (north) to lower (south) velocities. The emission extends ~120 kpc northwest of the center of 4C 41.17, which we identified as the radio core at right ascension 6^h 50^m 52^s.15, declination +41° 30′ 30″.8 (J2000 equinox) using additional observations with the Multi Element Remotely Linked Interferometer Network (MERLIN) telescope (Fig. 1) (*16*). One end of the [C I] emission is at the approximate location and velocity of 4C 41.17. The [C I] emission forms a coherent structure in both location and velocity, so we coadded the total signal using statistical weighting (*16*). We found that the integrated signal of the emission is inconsistent with the normal distribution of the noise background, with a total significance (neglecting systemic uncertainties) of 5.7σ (fig. S2). This statistical calculation (*16*) includes the regions labeled NW1 to NW4 in Fig. 1 but excludes the region W that is inside the Lyα halo.

Interpretation as a cold gas stream

We interpret the observed distribution and velocity of the [C I] emission as a stream of cold gas. The stream appears clumpy in Fig. 1, although this appearance is enhanced by the data sampling that we used to image the [C I] in the different regions. The emission extends for at least 100 kpc outside the Lyα halo, although the narrow-band imaging that we used to detect Lyα (Fig. 1) was restricted to velocities that do not overlap with those that we measured for the [C I] stream (*16*). The stream is aligned perpendicular to the major axis of both the Lyα halo and the radio jet. It appears to connect to a prominent dark lane in the Lyα emission, roughly 30 kpc west of the central radio galaxy (Fig. 1). We estimate (*16*) that the chance of randomly detecting a spurious streamlike signal of at least 5.7σ significance in our data is 0.08% (3.2σ). This calculation assumes that the stream is connected both spatially and kinematically to the [C I] emission in the central region of 4C 41.17 and accounts for confirmation bias (fig. S2) (*16*). We used an optimal stacking method to extract the integrated [C I] signal across the entire stream (Fig. 2), from which we calculated the stream's total mass of molecular hydrogen $M_{\rm H2}$ ~ (6.7 ± 2.2) × 10^{10} solar masses ($M_\odot$) (supplementary text). The inferred physical properties of the stream are listed in table S1.

Figure 3B shows a position velocity diagram of the [C I] emission that covers the stream and the central galaxy (Fig. 3A). The projected velocity of the stream steadily decreases at smaller offsets until it matches the Lyα gas kinematics in region W (*21*), where the stream connects to the 4C 41.17 system. This is consistent with gas accreting into the Lyα halo. In its outermost region (NW4), the stream has an initial velocity of v ~ 650 ± 150 km s^{-1} with respect to the gas at the base of the stream (NW1). This is larger than the typical velocities observed for tidal gas tails associated with galaxy interactions (*22*). The direction and kinematics of the stream appear inconsistent with an outflow, which would likely align along the jet axis and slow down going outward, contrary to the observed [C I] outside the Lyα halo. Stream velocities of ~650 km s^{-1} at distances of ≳100 kpc appear in simulations of gas accretion (*23*) for systems with dark matter halos of total mass ~3 × 10^{13} $M_\odot$, which is consistent with the value we derive for 4C 41.17 (supplementary text). The width of the stream is spatially unresolved, ≤30 kpc, which is not constraining but consistent with that of simulations (*5*). The [C I] in the central region of 4C 41.17 is shown in Fig. 4. This [C I] emission is spread across a velocity range of ~1000 km s^{-1}, which is similar to previous observations of carbon monoxide in the CO (J = 4→3) line (*24*), where J is the rotational quantum number. Because of the limited spatial resolution of our ALMA data, we cannot study the physical processes that occur in the gas in this central region. Figure 3B shows

[1]National Radio Astronomy Observatory, Charlottesville, VA 22903, USA. [2]Université Lyon 1, École Normale Supérieure de Lyon, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5574, Centre de Recherche Astrophysique de Lyon, F-69230 Saint-Genis-Laval, France. [3]Sorbonne Université, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7095, Institut d'Astrophysique de Paris, 75014 Paris, France. [4]Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel. [5]Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106, USA. [6]Department of Astronomy, Yale University, New Haven, CT 06511, USA. [7]Heidelberger Institut für Theoretische Studien, Schloss-Wolfsbrunnenweg 35, D-69118 Heidelberg, Germany. [8]Centro de Astrobiología, Consejo Superior de Investigaciones Científicas – Instituto Nacional de Técnica Aeroespacial, 28850 Torrejón de Ardoz, Madrid, Spain. [9]Sterrewacht, Leiden University, 2300 RA Leiden, Netherlands. [10]European Southern Observatory, 85748 Garching, Germany. [11]Instituto de Astrofísica de Andalucía, Consejo Superior de Investigaciones Científicas, E-18008 Granada, Spain. [12]Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, 50019 Zaragoza, Spain. [13]School of Sciences, European University Cyprus, Engomi, 1516 Nicosia, Cyprus. [14]School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK. [15]Institut Universitaire de France, Ministère de l'Enseignement Supérieur et de la Recherche, 75231 Paris, France.
***Corresponding author. Email: bemonts@nrao.edu**

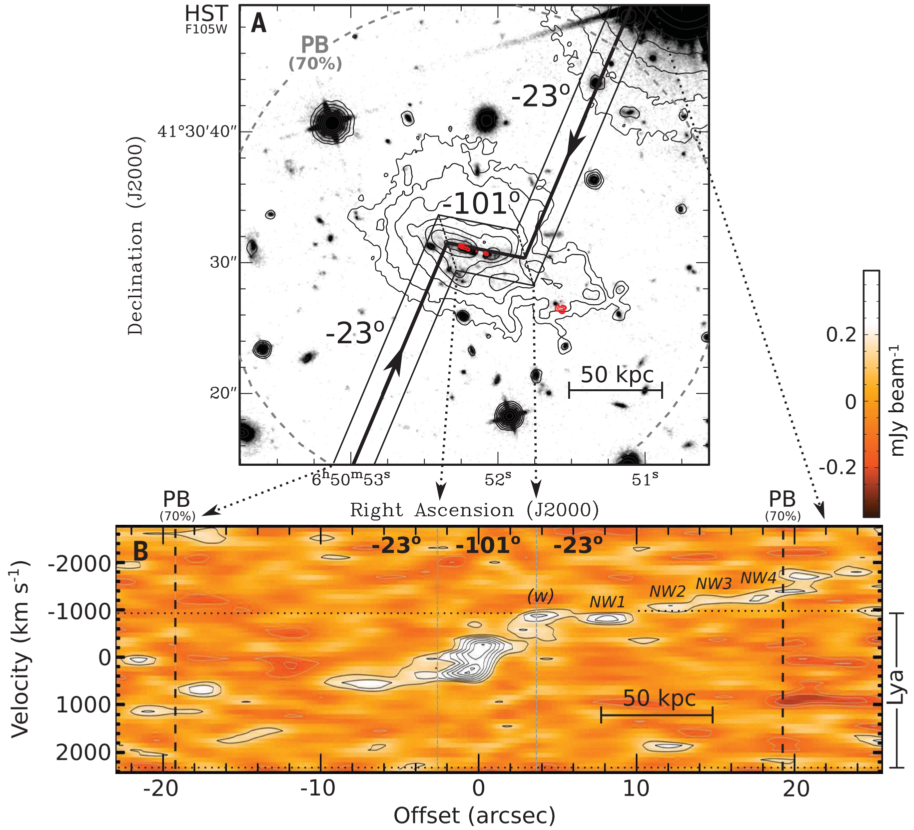


Fig. 3. Kinematics of the [C I]-emitting gas in the stream. (**A**) The background is an inverted grayscale image from the Hubble Space Telescope (HST) using the Wide Field Camera 3 (WFC3) and the F105W filter (*30*), which spans 900 to 1200 nm (188 to 250 nm in the rest-frame at z = 3.792). Black contours indicate the Lyα emission from recombining hydrogen gas (*17*). Red contours show the radio jet as in Fig. 1. The thick black line with arrows shows the direction used to derive the position-velocity diagram [as in (B)], with the thin black lines marking the boundaries of the included emission (equal to the size of the synthesized ALMA beam). The labeled position angles are measured from north to east. The dashed circle indicates the 70% response level of the PB. (**B**) Position-velocity diagram of the [C I] emission in the regions indicated in (A). Dotted arrows connect equivalent locations in (A) and (B). This assumes that the gas has the same redshift as that of the galaxy, z = 3.792. Black contours start at 130 μJy beam^{-1} and increase in steps of 65 μJy beam^{-1}. This corresponds to 2σ with steps of 1σ at the center of the galaxy (offset of zero). Negative contours are plotted at the same levels in gray. The data have been corrected for the PB response. The vertical dashed lines indicate the 70% response level of the PB, where the noise is 43% higher than at the center of the galaxy. The dotted horizontal lines mark the velocity extent of the Lyα data [(A) and Fig. 1]. Regions NW1, NW2, NW3, NW4, and W are the same as in Fig. 1.

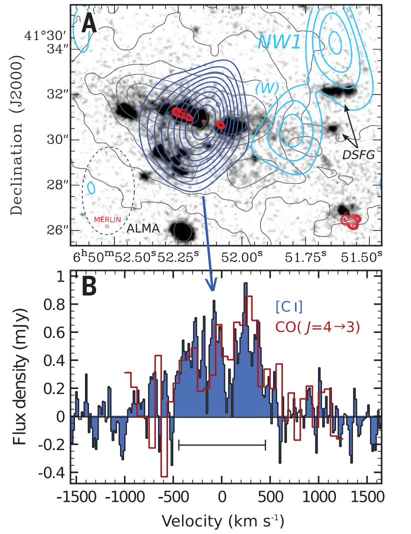


Fig. 4. Gas in the center of 4C 41.17. (**A**) Dark blue contours show the total intensity of [C I] emission in the center of 4C 41.17, overlain on the same HST image as Fig. 3A. Contour levels start at 2σ and increase in steps of 1σ, with σ = 0.038 Jy bm^{-1} km s^{-1}. The light blue contours show the [C I] emission from part of the stream, as in Fig. 1. The gray contours show the Lyα emission as in Fig. 1, whereas the red contours show the MERLIN data of the radio jet (*16*). Black arrows indicate two dusty star-forming galaxies (DSFGs) (fig. S4). (**B**) Hanning-smoothed spectrum of the [C I] emission in the center of 4C 41.17. Overlain in dark red is the spectrum of CO (J = 4⟶3) emission (*24*), scaled down by a factor of three in flux density for comparison. Both spectra are centered at redshift z = 3.792. The black bar indicates the velocity range −440 to +460 km s^{-1} that was used to integrate the [C I] emission to obtain the contours in (A).

additional emission southeast of 4C 41.17 in roughly the opposite direction to the stream (also visible in fig. S4).

Implications for the assembly of 4C 41.17

Because the stream contains carbon, it cannot consist of pristine gas from the Big Bang but must have been enriched with metals (elements heavier than helium). Cosmological simulations predict that low-mass galaxies and giant star-forming clumps flow along accreting gas streams (*2*, *13*, *25*, *26*), which could enrich the streams with metals because of outflows driven by star formation. Imaging of Lyα emission from cosmic filaments on larger scales has shown that most Lyα emission is powered by ultralow-luminosity dwarf galaxies (*9*). Streams and filaments might therefore contain a large number of dwarf galaxies, too faint to be visible in optical imaging. For 4C 41.17, two adjacent star-forming galaxies have previously been detected in near-infrared (K band) imaging (*27*), in a region that shows a deficit in Lyα emission (*17*). We found that they are close to, but not coincident with, the base of the stream (Fig. 4). Previous studies concluded that these are probably dusty star-forming galaxies that could absorb and scatter the Lyα emission (*17*, *27*). If these galaxies are at the same redshift as that of 4C 41.17, then they might be related to the stream; however, their redshifts have not been measured, so we cannot draw further conclusions from them. Southeast of these two star-forming galaxies is the prominent dark lane in the Lyα imaging where the stream appears to connect to 4C 41.17. At this dark lane, Lyα photons might be absorbed by dust or neutral hydrogen gas, potentially provided by the stream.

For the [C I] emission in the central region of 4C 41.17 (Fig. 4), we derive a total mass of molecular hydrogen gas of M_{H2} ~ (1.4 ± 0.2) × 10^{11} $M_\odot$ (supplementary text). The bulk of the stars in 4C 41.17 are part of an evolved stellar population, which has been forming stars at a continuously high rate of ≳250 $M_\odot$ year^{-1} for the previous 1 billion years (*28*). A recent burst of star formation, possibly enhanced locally by the radio jet (*29*), has temporarily raised the current star formation rate by an order of magnitude (*29*, *30*). Even a continuous star formation rate of ~250 $M_\odot$ year^{-1} would deplete the supply of molecular gas within 4C 41.17 in 560 million years, or before redshift z ~ 3. Assuming the stream is depositing molecular gas into 4C 41.17, we estimate its average accretion rate to be $\dot{M}_{acc}$ ~ 450 ± 180 $M_\odot$ year^{-1} (supplementary text). If the stream acquires

 https://avxhm.se/blogs/hill0

its gas from a larger reservoir, such as the cosmic web (*9*), then the stream could potentially supply sufficient gas to 4C 41.17 to sustain the high star formation rate for much longer than the current gas depletion time.

We conclude that the formation of 4C 41.17 is not an isolated process, but one linked to the surrounding intergalactic medium.

REFERENCES AND NOTES

1. D. Kereš, N. Katz, D. H. Weinberg, R. Davé, *Mon. Not. R. Astron. Soc.* **363**, 2–28 (2005).
2. A. Dekel *et al*., *Nature* **457**, 451–454 (2009).
3. F. van de Voort, J. Schaye, C. M. Booth, C. Dalla Vecchia, *Mon. Not. R. Astron. Soc.* **415**, 2782–2789 (2011).
4. J. Sánchez Almeida, B. G. Elmegreen, C. Muñoz-Tuñón, D. M. Elmegreen, *Astron. Astrophys. Rev.* **22**, 71–131 (2014).
5. N. Mandelker *et al*., *Mon. Not. R. Astron. Soc.* **494**, 2641–2663 (2020).
6. D. C. Martin *et al*., *Nature* **524**, 192–195 (2015).
7. H. Umehata *et al*., *Science* **366**, 97–100 (2019).
8. E. Daddi *et al*., *Astron. Astrophys.* **649**, A78 (2021).
9. R. Bacon *et al*., *Astron. Astrophys.* **647**, A107 (2021).
10. M. Dijkstra, A. Loeb, *Mon. Not. R. Astron. Soc.* **400**, 1109–1120 (2009).
11. C.-A. Faucher-Giguère, D. Kereš, M. Dijkstra, L. Hernquist, M. Zaldarriaga, *Astrophys. J.* **725**, 633–657 (2010).
12. J. Rosdahl, J. Blaizot, *Mon. Not. R. Astron. Soc.* **423**, 344–366 (2012).
13. D. Ceverino, A. Dekel, F. Bournaud, *Mon. Not. R. Astron. Soc.* **404**, 2151–2169 (2010).
14. K. C. Chambers, G. K. Miley, W. J. M. van Breugel, *Astrophys. J.* **363**, 21–39 (1990).
15. G. K. Miley, K. C. Chambers, W. J. M. van Breugel, F. Macchetto, *Astrophys. J.* **401**, L69–L73 (1992).
16. Materials and methods are available as supplementary materials.
17. M. Reuland *et al*., *Astrophys. J.* **592**, 755–766 (2003).
18. P. P. Papadopoulos, W.-F. Thi, S. Viti, *Mon. Not. R. Astron. Soc.* **351**, 147–160 (2004).
19. T. G. Bisbas *et al*., *Astrophys. J.* **839**, 90–107 (2017).
20. S. C. O. Glover, P. C. Clark, *Mon. Not. R. Astron. Soc.* **456**, 3596–3609 (2016).
21. M. Reuland *et al*., *Astron. J.* **133**, 2607–2623 (2007).
22. P. Serra *et al*., *Mon. Not. R. Astron. Soc.* **428**, 370–380 (2013).
23. T. Goerdt, D. Ceverino, *Mon. Not. R. Astron. Soc.* **450**, 3359–3370 (2015).
24. C. De Breuck *et al*., *Astron. Astrophys.* **430**, L1–L4 (2005).
25. N. Mandelker, P. G. van Dokkum, J. P. Brodie, F. C. van den Bosch, D. Ceverino, *Astrophys. J.* **861**, 148–169 (2018).
26. M. Fumagalli *et al*., *Mon. Not. R. Astron. Soc.* **418**, 1796–1821 (2011).
27. J. R. Graham *et al*., *Astrophys. J.* **420**, L5–L8 (1994).
28. B. Rocca-Volmerange *et al*., *Mon. Not. R. Astron. Soc.* **429**, 2780–2790 (2013).
29. A. Dey, W. van Breugel, W. D. Vacca, R. Antonucci, *Astrophys. J.* **490**, 698–709 (1997).
30. G. Drouart *et al*., *Astron. Astrophys.* **593**, A109 (2016).

ACKNOWLEDGMENTS

The authors thank G. Drouart for useful discussions. This paper makes use of the following ALMA data: ADS/JAO. ALMA#2018.1.01334.S. ALMA is a partnership of ESO (representing its member states), NSF (USA), and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO, and NAOJ. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. MERLIN is a National Facility operated by the University of Manchester at Jodrell Bank Observatory on behalf of STFC. This paper uses observations made with the NASA/ESA Hubble Space Telescope, obtained from the data archive at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. **Funding:** Support for program number HST-AR-16123.001-A was provided through a grant from the STScI under the National Aeronautics and Space Administration (NASA) contract NAS5-26555. The MERLIN observations benefited from research funding from the European Community's sixth Framework Programme under RadioNet R113CT 2003 5058187. M.D.L. thanks the Centre National de la Recherche Scientifique/Institut National des Sciences de l'Univers (CNRS/INSU) for their financial and administrative support. N.M. acknowledges partial support from the Gordon and Betty Moore Foundation through grant GBMF7392, from the National Science Foundation through grant NSF PHY-1748958, and from the Klauss Tschira Foundation through the Heidelberg Institute for Theoretical Studies (HITS) Yale Program in Astrophysics (HYPA). M.V.-M. acknowledges support through grant PID2021-124665NB-I00 funded by the Spanish Ministry of Science and Innovation (MCIN)/State Agency of Research (AEI)/10.13039/501100011033 and by the European Regional Development Fund (ERDF) "A Way of Making Europe." M.A.P.-T. acknowledges financial support through grants CEX2021-001131-S and PID2020-117404GB-C21 funded by MCIN/AEI/10.13039/501100011033. N.A.H. acknowledges support from the Science and Technology Facilities Council (STFC) consolidated grant ST/T000171/1. **Author contributions:** B.H.C.E. proposed the ALMA observations, processed and analyzed the ALMA data, interpreted the results, and wrote the manuscript. M.D.L. verified the data analysis, interpreted the results, and wrote the manuscript. I.Y. processed the ALMA data and interpreted the results. N.M. performed the comparison with simulations and contributed to manuscript preparation. M.V.-M. provided the comparison with Lyα and interpreted the results. G.K.M. obtained the HST and Lyα imaging and discussed the results. C.D.B. provided the MERLIN data and analyzed the host galaxy. M.A.P.-T. processed the MERLIN data and contributed to manuscript preparation. N.A.H. contributed to the HST and Lyα imaging. P.G. contributed to the interpretation. All authors also discussed the results and provided feedback on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The ALMA data are archived at https://almascience.nrao.edu/aq/ under project code 2018.1.01334.S. The HST data are available at https://hla.stsci.edu/hlaview.html under advanced search proposal ID 11738. The MERLIN data are provided as data S1. The code we used to assess the statistical significance is archived at https://github.com/bjornemonts/stream-statistics/releases

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abh2150
Materials and Methods
Supplementary Text
Figs. S1 to S5
Table S1
Movie S1
Data S1
References *(31–93)*

Submitted 5 March 2021; accepted 14 February 2023
10.1126/science.abh2150

Check for updates

APPLIED PHYSICS

Fiber pumps for wearable fluidic systems

Michael Smith[1]*, Vito Cacucciolo[1,2], Herbert Shea[1]*

Incorporating pressurized fluidic circuits into textiles can enable muscular support, thermoregulation, and haptic feedback in a convenient wearable form factor. However, conventional rigid pumps, with their associated noise and vibration, are unsuitable for most wearables. We report fluidic pumps in the form of stretchable fibers. This allows pressure sources to be integrated directly into textiles, enabling untethered wearable fluidics. Our pumps consist of continuous helical electrodes embedded within the walls of thin elastomer tubing and generate pressure silently through charge-injection electrohydrodynamics. Each meter of fiber generates 100 kilopascals of pressure, and flow rates approaching 55 milliliters per minute are possible, which is equivalent to a power density of 15 watts per kilogram. The benefits in design freedom are considerable, which we illustrate with demonstrations of wearable haptics, mechanically active fabrics, and thermoregulatory textiles.

The widespread use of textiles and their close contact with the human body make functional fabrics an attractive technology for wearable devices. When creating technology that is meant to be worn, fiber-format components offer several advantages because of their flexibility, scalability, and compatibility with existing textile production techniques (*1*, *2*). Fiber-based actuators (*3*), sensors (*4*), energy storage (*5*), and energy harvesting (*6*) devices demonstrate the progress in this field and the benefits afforded by this format.

Lacking, however, are fiber-based fluidic components. Fluidic actuation is fundamental to much of soft robotics—a key axis of wearable technology (*7*, *8*). Fluidic systems also enable numerous other processes, including heat and mass transport, within conventional medical, industrial, and personal devices. The lack of active fluidic components that can be straightforwardly integrated into textiles is inhibiting the development of countless useful devices—devices such as soft supportive exoskeletons, adaptive medical orthoses, and wearable haptics.

Generating meaningful fluidic power (the product of pressure and flowrate) in a wearable and portable manner remains a substantial challenge. Although several "soft" pumps made entirely from compliant materials have recently been reported (*9*–*13*), their low levels of power density hinder untethered wearable applications. Furthermore, none of these pumps has a form factor that can be woven into a textile, and none demonstrate the scalability necessary for human-sized applications.

We present a fluidic pump in the form of a fiber (Fig. 1A). The pump is a flexible and stretchable tube, on the order of millimeters in diameter, that can generate continuous fluid flow without any moving parts or vibration. The amount of pressure and flowrate generated by the fibers is a function of both their length (Fig. 1E) and diameter (Fig. 2B).

The fiber pumps are able to pump liquid in a form factor that is inherently compatible with wearable devices (Fig. 1F). In generating flow directly within the tubing itself (Fig. 1G), the need for an external pump is removed, and fluidic systems can be greatly simplified. This affords substantial improvements in design freedom for wearables. Pumps can be distributed throughout the volume of the device, reducing losses, improving comfort, and enabling advanced multipump wearables without the need for valves and connectors. The distributed nature of the fiber pumps means that the weight of the pump can be counted against the weight of tubing required in a conventional fluidic circuit. In the case of fiber pumps, the tubing is the pump, and effectively the pump comes at no extra weight penalty.

Fiber pump fabrication and performance

The pumps operate through the principle of charge injection electrohydrodynamics (EHD) (*9*, *14*–*16*). Embedded in the pumps' polyurethane tube wall are two continuous helical electrodes, fabricated from copper wire (80 μm diameter). Applying a dc electric field up to 8 kV/mm between these electrodes ionizes a dielectric liquid within the tube, creating negatively charged ions as liquid molecules accept electrons. These ions are accelerated toward the positive electrode, where they discharge. As they move, they set in motion the surrounding liquid molecules to create a net fluid flow (Fig. 1C). The asymmetrical spacing of the helical electrodes ensures that for a given polarity of applied voltage, there is a net flow in one direction. Inverting the polarity of the applied voltage will reverse the direction of flow (fig. S7A).

Helical configurations of electrodes have previously been simulated for use in EHD conduction pumps (a related but distinct pumping mechanism) (*17*). We demonstrate that this configuration can produce exceptional performance as an ion-drag EHD pump. We report pressures exceeding 80 kPa and flow rates approaching 55 ml/min. Considering the weight of the pump, this corresponds to power densities in excess of 15 W/kg, which is substantially larger than previous attempts at soft pumps and within a factor of two of class-leading (rigid) miniature pumps (table S3). Furthermore, this configuration is up to 10 times more efficient than a conventional layout of interdigitated electrodes (*9*); can generate 30 times the pressure response rate (Fig. 2F); and results in a stable, repeatable device (fig. S6).

The structure of the fiber pumps is created by using a filament winding fabrication method (Fig. 1D, fig. S5, and movie S2). This involves simultaneously twisting thermoplastic polyurethane (TPU) filament and copper wire together around a central mandrel and subsequently fusing the filament together with heat. Removing the mandrel reveals a hollow tube with helical electrodes embedded within the walls (*18*). Crucially, these electrodes remain exposed along the inner surface of the tube wall (Fig. 1B, ii), enabling the injection and collection of charge necessary for EHD pumping. The filament winding method is continuous and allows for a simple pump construction from only two materials. This greatly reduces the compatibility issues between pump materials and EHD liquids that were previously reported (*9*).

This same fabrication method can be used to create different pump geometries. The exact structure is determined by the close packing of the TPU filaments around the mandrel, which is dependent on the number of filaments, their diameter, and the diameter of the mandrel (supplementary text and fig. S1). Changing the mandrel diameter leads to different diameter pumps (Fig. 2A). Because the structure is helical, however, the pump diameter cannot be changed independently; some other aspect of the structure must also change. For the pumps shown in Fig. 2, the helix angle has been kept constant at ~60° by changing the number of filaments for each diameter (table S1). Consequently, the helix pitch is different in each case.

Under these conditions, reducing the pump diameter increases the maximum pressure but decreases the maximum flowrate (Fig. 2B, i). A maximum power density of 15.2 W/kg occurs at a diameter of 2.3 mm (1.5 mm inner diameter) (Fig. 2B, ii), although the optimum helix angle and pitch are likely different for each diameter (supplementary text and fig. S2). The electrode spacing can be controlled independently

[1]Soft Transducers Laboratory (LMTS), École Polytechnique Fédérale de Lausanne, Neuchâtel, Switzerland. [2]Department of Mechanics, Mathematics and Management (DMMM), Politecnico di Bari, Bari, Italy.
*Corresponding author. Email: michael.smith@epfl.ch (M.S.); herbert.shea@epfl.ch (H.S.)

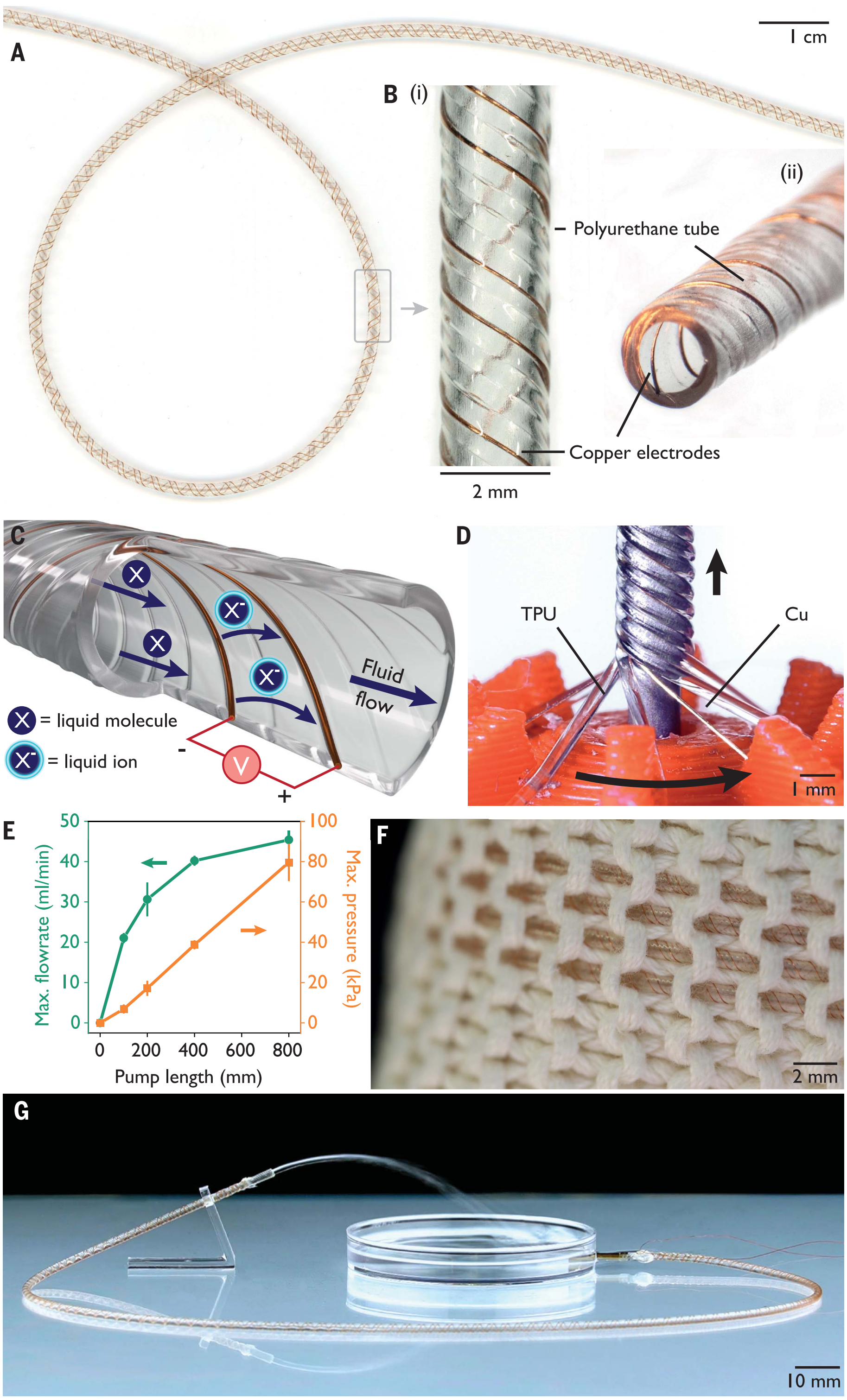


Fig. 1. A fiber-format electrohydrodynamic pump. (**A**) The fiber pump is soft and flexible. (**B**) (i) The structure of the fiber pump, showing (ii) the helical copper electrodes embedded within the walls of a polyurethane tube. (**C**) A schematic illustration of the pump operating principle. Liquid molecules become ionized at the negative electrode, accepting an electron. The ions are then accelerated to and discharged at the positive electrode, generating a net fluid flow. (**D**) Pumps are fabricated by using a filament-winding method. (**E**) The scaling of fiber pump pressure and flow rate with pump length, for an applied electric field of 8 kV/mm. Values are mean ± SD for n = 3 pumps of each length. (**F**) The fiber pumps sewn into a woolen textile. (**G**) The fiber pump in operation (movie S1).

of diameter and is set by the number of TPU filaments between the copper wires. In all cases in this study, two filaments were used, giving a (perpendicular) electrode spacing of 800 μm. We anticipate that the pump structure can be optimized further, leading to additional improvements in performance.

With respect to the applied electric field, the maximum pressure scales nonlinearly (Fig. 2C), as is typically seen in charge-injection EHD pumps (*15*). Maximum flowrate scales linearly with applied field once above a value of 5 ml/min (Fig. 2D), a behavior also seen in other EHD pumps (*15*, *19*). Peak power density and efficiency both increase with increasing pump length (Fig. 2E) and electric field (fig. S7, E and F). The fiber pumps can pump continuously for up to 6 days before chemical deposits passivate the electrodes (fig. S7H). This is considerably longer than any other reported EHD pump, soft or otherwise (*9*), and can likely be improved further through careful selection of liquid and electrode material.

Throughout this work, we used the dielectric liquid Novec 7100 (3M). This is a nontoxic, nonflammable methoxy-fluorocarbon with low global warming potential (*20*). It is commonly used as a solvent and for heat management but also performs well as an EHD fluid because of readily ionizable fluorine groups and a high electrical breakdown field of 10 kV/mm (*20*). Other liquids (including nonfluorinated compounds) may be pumped, provided that they have low conductivity and viscosity and are electrochemically stable (*21*, *22*).

The electric fields used to drive these pumps equate to voltages on the order of kilovolts. This is necessary to overcome the energy barrier for ionization of the liquid. However, the amount of current flowing is low (<110 μA) (fig. S7I). This results in a low power consumption of ~0.9 W/m. Consequently, the fiber pumps can be powered with a battery-operated high-voltage power supply, weighing 35 g [including battery (fig. S11)], enabling untethered operation (movie S1). A typical smartphone battery would power a 1-m length of pump continuously for ~15 hours.

Pump flexibility and stretchability

A major attraction of fiber-based technologies is the compliance offered by this format. The fiber pumps are flexible about every direction perpendicular to their axis, offering substantial design freedom for integration into wearables and highly compliant soft actuators. Therefore, we have characterized how the fiber pumps behave when deformed. Stretchability was assessed by measuring the pressure-flowrate characteristics of the pump—known as the "pump curve"—at different values of uniaxial strain (Fig. 2G). Even at 15% strain (a strain level sufficient for wearable applications),

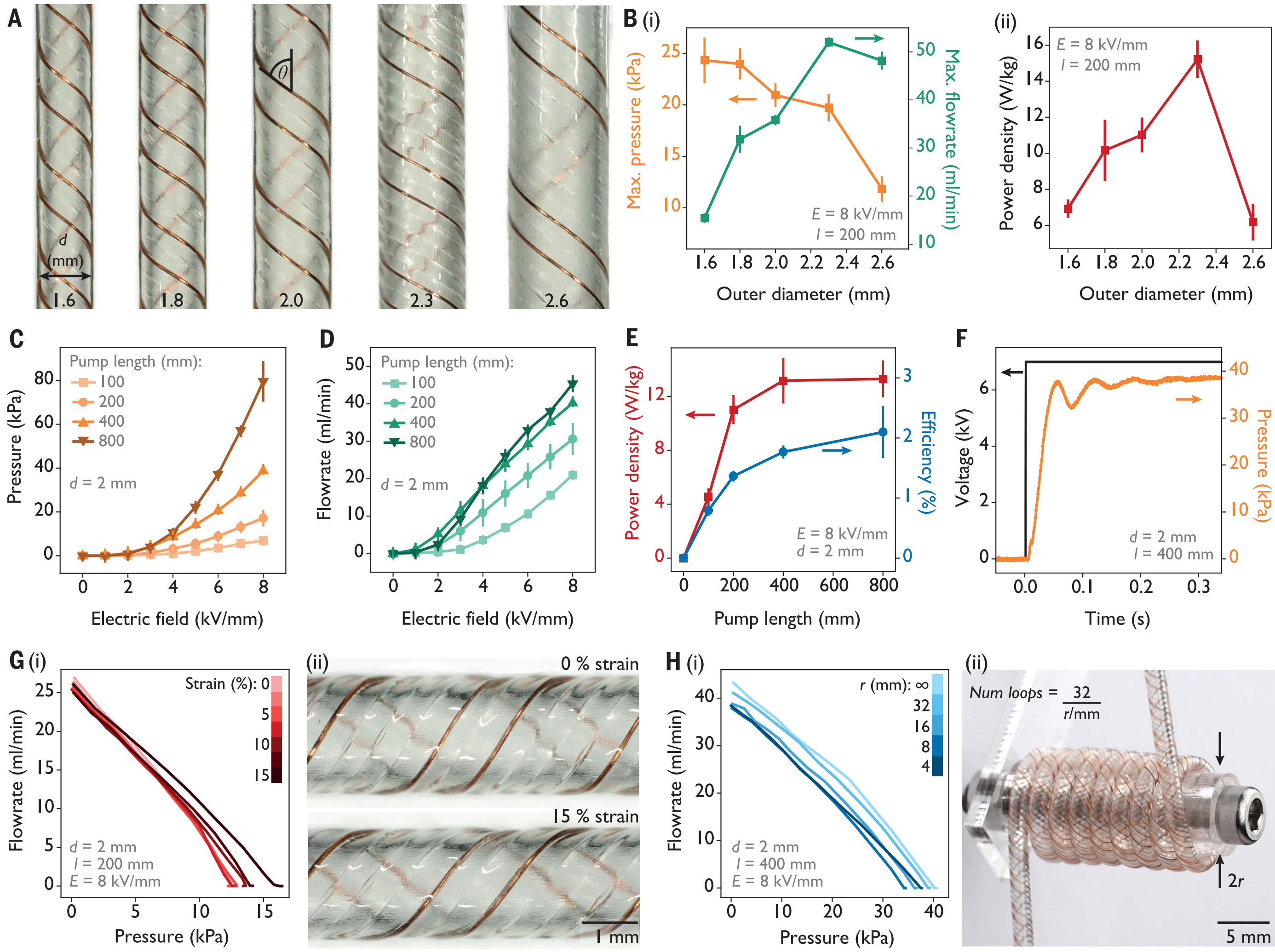


Fig. 2. Characterization of the fiber pumps. Values with error bars correspond to mean ± SD for n = 3 pumps. (**A**) Fiber pumps fabricated at different diameters (d) with an approximately constant helix angle θ ~ 60°. (**B**) Variation in (i) maximum pressure and flowrate and (ii) power density for the different pump diameters at a constant helix angle of 60°. l, length; E, applied electric field. (**C**) Pressure and (**D**) flowrate scaling with applied electric field for various pump lengths, all with an outer diameter of 2 mm. (**E**) The power density and efficiency as a function of pump length, for a 2-mm-diameter pump at 8 kV/mm. (**F**) Pressure and voltage as a function of time. Maximum stable pressure is reached within 200 ms. The oscillations in the pressure measurement are due to the elasticity of the pump and connecting tubing. (**G**) (i) The pressure-flowrate characteristics of a fiber pump at (ii) different levels of uniaxial strain. (**H**) (i) The pressure-flowrate characteristics of a fiber pump subjected to different radii of curvature (r). Curvature is achieved by wrapping the pump around circular objects with multiple loops (ii) to ensure that an equal length of pump was deformed in each case.

pump performance was broadly unchanged. Beyond 25% strain, the pump would begin to fail as the electrode structure detached from the tube wall (fig. S8). This stretchability was achieved by using conventional materials: The helical nature of the electrodes means that a stretchable device can be fabricated without lossy and unreliable stretchable conductive materials (Fig. 2G, ii).

The pump curves captured as the fibers are deformed to different radii of curvature are displayed in Fig. 2H, i. To enable a fair comparison between different curvatures, we adjusted the number of loops to ensure that an equal length of pump was deformed in each case (Fig. 2H, ii). Deforming to a bend radius of 4 mm—four times the radius of the pump itself—reduced the maximum pressure by 7% and maximum flowrate by 12% when compared with an undeformed pump. The slight reduction can be explained by considering the change in electrode spacing around the inside and outside of the deformed pump, as well as the pressure loss inherent to fluid flow through a curved tube. The pump will kink if subjected to a sufficiently low bend radius, typically <3 mm, although the exact value will depend on the loading conditions and pump dimensions (*23*). The kink, however, does not damage the pump structure, and the pump can continue to operate once the kink is removed (fig. S8 and movie S9).

Under repeated uniaxial strain, the pumps can withstand at least 5000 cycles to 10% strain without change in performance (fig. S9). Repeated flexion to a bend radius of 3 mm for 5000 cycles has no effect on pump performance (fig. S10).

Fiber and textile-based actuators with fiber pumps

The ability of these pumps to generate substantial flow rates and pressures while maintaining flexibility permits multiple applications in textile-based and wearable applications, which previously required an external conventional pump. We have demonstrated the fiber pumps in the context of mechanically active fibers

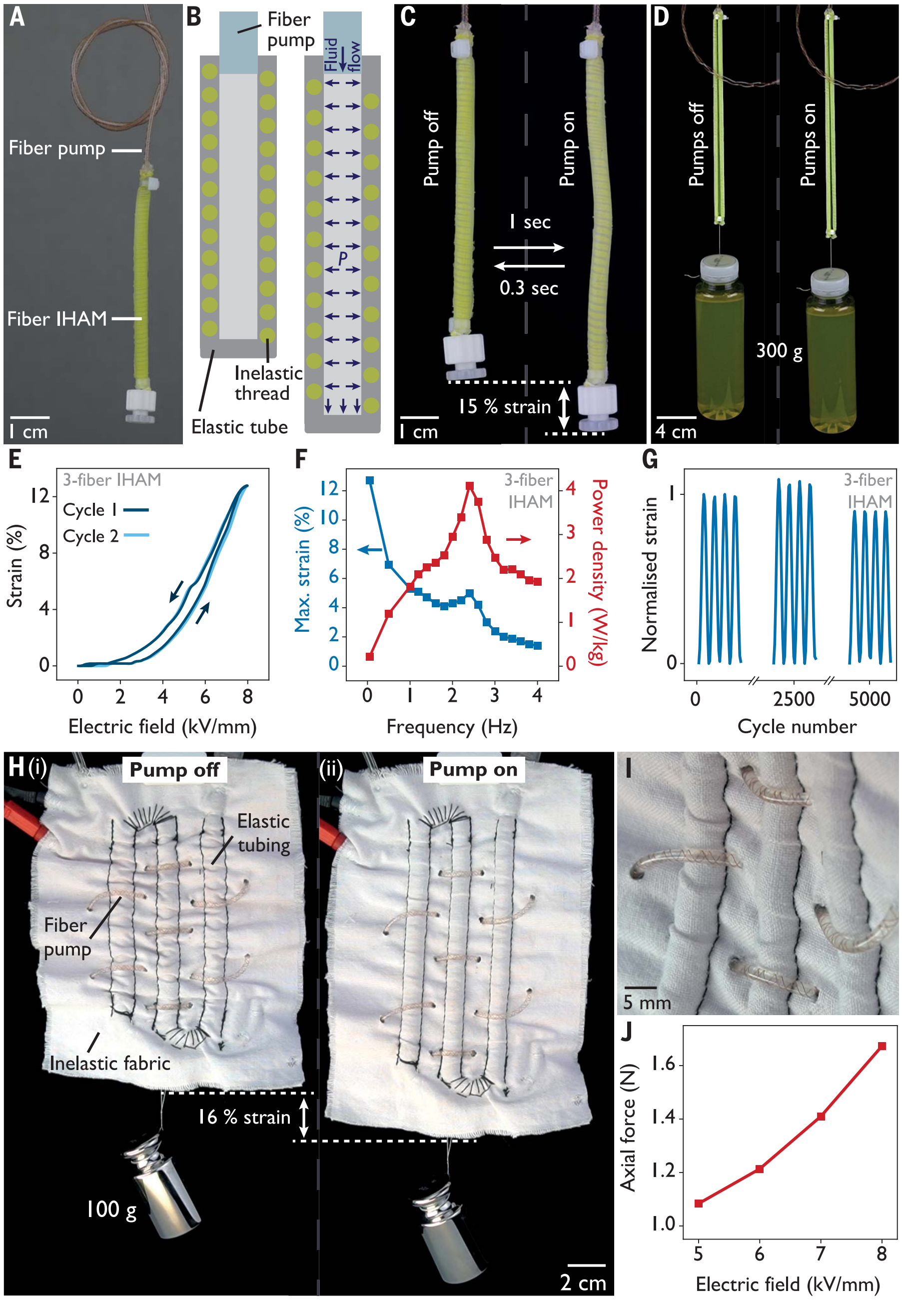


Fig. 3. Actuators powered using fiber pumps. (**A**) A fiber pump connected to a 9-cm-long inverse hydraulic artificial muscle (IHAM). (**B**) The IHAM consists of helices of inelastic thread embedded within the wall of an elastic tube. (**C**) A maximum strain of 15% is achieved from a single fiber. (**D**) Combining IHAMs and pumps in parallel can generate higher forces. In this image, three IHAMs and pumps are used to move a 300 g load. (**E**) The strain of the three-fiber IHAM as a function of the electric field applied to the fiber pumps. (**F**) The frequency response of the three-fiber IHAM moving a 200-g weight. Applied signal is 6.4 kV square wave (equivalent to 8 kV/mm). We observed a small resonance peak at ~2.4 Hz. (**G**) The three-fiber IHAM can operate for at least 5000 cycles at 2.4 Hz, with minimal loss in amplitude. (**H**) A fluidic fabric actuator with integrated fiber pump. In this device, an elastic tube is sewn between two sheets of inextensible fabric. The ruffling of the fabric allows the actuator to extend when pressurized. (**I**) The fiber pump integrated directly into the fabric actuator. (**J**) The maximum difference in isometric axial force generated by the actuator as a function of the applied electric field.

and textiles (Fig. 3), as well as thermoregulatory wearables for heat management and haptic feedback (Fig. 4).

Actuation can be achieved by using fiber-like artificial muscles, which are similar in concept to inverse pneumatic artificial muscles (IPAMs) (*24*) but operate with a liquid rather than a gas. These inverse hydraulic artificial muscles (IHAMs) can be integrated directly with the fiber pumps (Fig. 3A and movie S8). The IHAM consists of an elastic tube with a helix of inelastic thread embedded within its walls. This thread acts as a radial constraint so that as the IHAM is pressurized by the fiber pump, the device does not expand radially but instead extends along its length (Fig. 3B). Actuation is fast (Fig. 3C) because of the low internal volume of the actuator (~ 300 μl) and the high performance of the fiber pump. A single fiber exerts a maximum difference in isometric axial force of ~0.5 N, and actuators can be used in parallel to increase force output (Fig. 3D). Response to the applied voltage is repeatable, with low hysteresis (Fig. 3E). The fast time response of the fiber IHAMs facilitates operation beyond the quasi-static regime (Fig. 3F), where they can continue to operate for several thousands of cycles (Fig. 3G).

Fiber pumps can be integrated directly into textiles to create active fabrics, as demonstrated in Fig. 3H. The device shown consists of a fluidic fabric muscle (*25*) with a fiber pump woven directly into the fabric (Fig. 3I). The operating principle is similar to that of the IHAMs, except here the radial constraint is supplied by an inextensible fabric sewn around the elastic tubing. The result is an active fabric, capable of changing shape upon electrical command (movie S7). Operating at an applied field of 8 kV/mm, maximum strains approaching 40% are possible (fig. S12D), and actuation strains of 16% were reached with an applied load of 100 g (Fig. 3H). We achieved a maximum isometric axial force of 1.7 N, which was controllable by tuning the pump voltage (Fig. 3J). Actuation speed was reduced because of the larger internal volume; nonetheless, full actuation occurred in under 20 s, an adequate timeframe for wearable applications such as self-adjusting clothing and intermittent compression devices. The entire device can also be washed with standard laundry detergent (fig. S13 and movie S10) without compromising its performance.

The example actuators shown in Fig. 3 use the pump's working fluid directly. It is equally feasible to instead use this fluid to displace a second fluid, such as air, to allow existing pneumatic actuators to be used without modification.

Thermoregulation and thermal haptics with fiber pumps

We also demonstrated two applications of fiber pumps in thermally active textiles. Shown in Fig. 4A is a thermal haptic glove, a class of device used to generate different temperature sensations to enhance the sense of immersion in virtual reality (*26*). Typically reliant upon rigid thermoelectric ceramics, in this work we demonstrated how a thermal haptic glove can be realized by using soft and flexible fiber pumps.

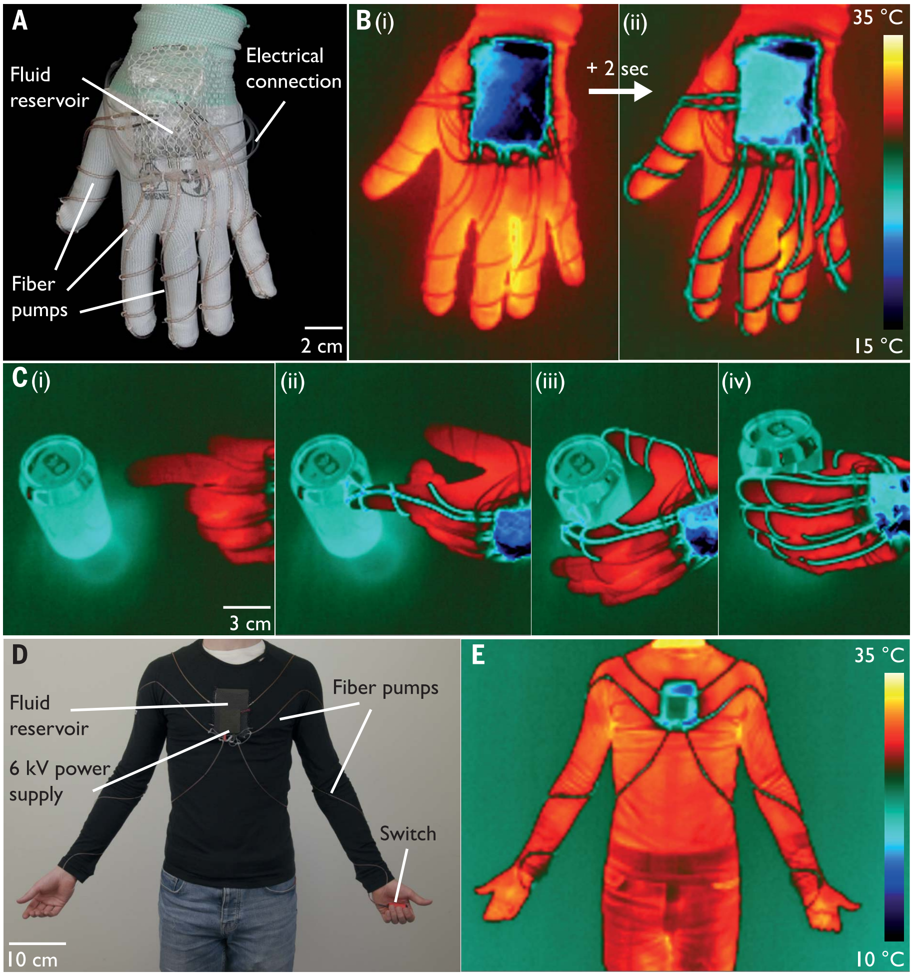


Fig. 4. Heat management by use of fiber pumps. (**A**) A thermal haptic glove with separate, individually controlled fiber pumps attached to each finger. (**B**) Thermal images of the glove (i) before and (ii) 2 s after simultaneously activating all pumps. (**C**) A demonstration of thermal haptic feedback (i) approaching the object with no pumps active, (ii) activating the pump of the finger in contact with the object, chilling the finger and providing a thermal haptic stimulus, subsequently activating (iii) the thumb and (iv) remaining fingers as they come into contact with the object. (**D**) A fully wearable and untethered thermoregulatory garment, with integrated fiber pumps, power supply and fluid reservoir.

The glove consists of a reservoir of chilled Novec 7100 fluid and five separate, independently controlled fiber pumps winding around the fingers. Activating a pump rapidly flows cold fluid around the finger (Fig. 4B), chilling the pump to the fluid temperature within 2 s and providing a thermal haptic stimulus to the user (movie S4). Once the pump voltage is removed and without the constant flow of chilled fluid from the reservoir, the pump returns to room temperature. In a virtual reality setting, this allows the temperature and thermal properties of objects to be simulated with a degree of spatiotemporal resolution: The physical object that represents the virtual one can be made to feel cold, and that sensation is felt only through those fingers in contact with the object (Fig. 4C). Achieving the same effect with any other pump would require several separate pumps or an array of valves, hindering wearability.

The same principle can be applied to an entire garment (Fig. 4D). Four sections of fiber pump transport chilled fluid around the torso and arms. The pumps are powered with a battery-operated power supply (fig. S11), creating an untethered system that can be used in any environment and during physical activity (movie S5). As well as enabling untethered operation, the low power consumption of the pumps means that Joule heating is negligible in comparison with the cooling rate provided by the chilled fluid.

The concept of thermoregulatory fabric can be developed further (fig. S14). The pump fibers themselves can be woven together to create a fluidic fabric that, when attached to hot and cold reservoirs of fluid, allows the temperature of the fabric to be regulated (movie S6). Such a patch of woven pumps can also be considered as a pressure source, sewn directly into the textile of a wearable device to provide high-pressure fluid for actuation.

Concluding remarks

We have demonstrated a fluidic pump in the form of a flexible fiber that is capable of generating exceptional pressures and flow rates. This advancement allows rigid and bulky external pumps to be removed from wearable fluidic devices—effectively turning the tubing into the pump. The scalability, simplicity, and stability of both the operation and fabrication of these fibers enable applications that are not possible with conventional pumping technologies.

REFERENCES AND NOTES

1. G. Loke, W. Yan, T. Khudiyev, G. Noel, Y. Fink, *Adv. Mater.* **32**, e1904911 (2020).
2. J. Xiong, J. Chen, P. S. Lee, *Adv. Mater.* **33**, e2002640 (2021).
3. C. S. Haines *et al*., *Science* **343**, 868–872 (2014).
4. W. Yan *et al*., *Nature* **603**, 616–623 (2022).
5. T. Khudiyev *et al*., *Mater. Today* **52**, 80–89 (2022).
6. H. Sun, Y. Zhang, J. Zhang, X. Sun, H. Peng, *Nat. Rev. Mater.* **2**, 17023 (2017).
7. T. Abe, S. Koizumi, H. Nabae, G. Endo, K. Suzumori, in *2018 IEEE International Conference on Soft Robotics (RoboSoft)* (2018), pp. 572–578.
8. O. Kilic Afsar *et al*., in *The 34th Annual ACM Symposium on User Interface Software and Technology* (Association for Computing Machinery, New York, NY, USA, 2021), pp. 1010–1026.
9. V. Cacucciolo *et al*., *Nature* **572**, 516–519 (2019).
10. Y. Matia, H. S. An, R. F. Shepherd, N. Lazarus, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2203116119 (2022).
11. R. S. Diteesawat, T. Helps, M. Taghavi, J. Rossiter, *Sci. Robot.* **6**, eabc3721 (2021).
12. W. Tang *et al*., *Nat. Commun.* **12**, 2247 (2021).
13. C. Stergiopulos *et al*., in *ASME 2014 Conference on Smart Materials, Adaptive Structures and Intelligent Systems* (American Society of Mechanical Engineers Digital Collection, 2014); https://asmedigitalcollection.asme.org/SMASIS/proceedings/SMASIS2014/46155/V002T04A011/286215.
14. A. Ramos, Ed., *Electrokinetics and Electrohydrodynamics in Microsystems* (Springer Vienna, 2011).
15. J. Kim, Y. Yamada, S. Yokota, *Int. J. Adv. Manuf. Technol.* **106**, 627–639 (2020).
16. Z. Mao, T. Iizuka, S. Maeda, *Sens. Actuators A Phys.* **332**, 113168 (2021).
17. N. O'Connor, J. Yagoobi, in *2021 IEEE Industry Applications Society Annual Meeting (IAS)* (2021); https://ieeexplore.ieee.org/document/9677332.
18. A detailed account of pump fabrication is provided in the materials and methods section of the supplementary materials (fig. S5 and movie S2).
19. T. Matsubara, H. H. Huynh, K. Yoshida, J. Kim, *Sens. Actuators A Phys.* **295**, 317–323 (2019).
20. M. S. EL-Genk, H. Bostanci, *Int. J. Heat Mass Transf.* **46**, 1841–1854 (2003).
21. S. Yokota, Y. Otsubo, K. Edamura, Electro-sensitive movable fluids, methods of using the same and motors for the electro-sensitive movable fluids (2000); https://patents.google.com/patent/US6030544A/en.
22. J. E. Bryan, J. Seyed-Yagoobi, *IEEE Trans. Electr. Insul.* **26**, 647–655 (1991).
23. K. Luo, P. Rothemund, G. M. Whitesides, Z. Suo, *J. Mech. Phys. Solids* **131**, 230–239 (2019).
24. E. W. Hawkes, D. L. Christensen, A. M. Okamura, in *2016 IEEE International Conference on Robotics and Automation (ICRA)* (2016), pp. 4022–4029.
25. M. Zhu, T. N. Do, E. Hawkes, Y. Visell, *Soft Robot.* **7**, 179–197 (2020).
26. Y. Huang *et al*., *Mater. Today Phys.* **22**, 100602 (2022).
27. M. Smith, V. Cacucciolo, H. Shea, Research data supporting 'Fiber pumps for wearable fluidic systems'. Zenodo (2023); https://doi.org/10.5281/zenodo.7451722.

https://avxhm.se/blogs/hill0

ACKNOWLEDGMENTS

The authors thank V. Py and M. Benbedda for assistance with the high-voltage electronics featured in this work. We thank S. Maeda and F. Hartmann for helpful discussions. We also thank J. La Scala for developing the automated fabrication system used in this work. Thanks also to L. Caillaud and E. Valicka for help in creating the textile-based demonstrations. Further thanks to E. Valicka and R. Hennig for assistance with videography and video editing. **Funding:** We acknowledge the support of the Swiss National Science Foundation (grant IZLJZ2_183656) **Author contributions:** M.S., V.C., and H.S. formulated the concept; M.S. fabricated the devices, carried out the experiments, analyzed the data, and created the figures and videos, with supervision from H.S. and V.C; M.S. prepared the original draft of the manuscript; M.S., V.C., and H.S. reviewed and edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data that support the claims in this manuscript are available on the Zenodo repository (*27*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade8654
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S3
Movies S1 to S10

Submitted 23 September 2022; accepted 28 February 2023
10.1126/science.ade8654

Check for updates

SOIL RESPIRATION

Soil respiration–driven CO_2 pulses dominate Australia's flux variability

Eva-Marie Metz[1]*, Sanam N. Vardag[1,2], Sourish Basu[3,4], Martin Jung[5], Bernhard Ahrens[5], Tarek El-Madany[5], Stephen Sitch[6], Vivek K. Arora[7], Peter R. Briggs[8], Pierre Friedlingstein[9,10], Daniel S. Goll[11], Atul K. Jain[12], Etsushi Kato[13], Danica Lombardozzi[14], Julia E. M. S. Nabel[5,15], Benjamin Poulter[16], Roland Séférian[17], Hanqin Tian[18], Andrew Wiltshire[19], Wenping Yuan[20], Xu Yue[21], Sönke Zaehle[5], Nicholas M. Deutscher[22], David W. T. Griffith[22], André Butz[1,2,23]*

The Australian continent contributes substantially to the year-to-year variability of the global terrestrial carbon dioxide (CO_2) sink. However, the scarcity of in situ observations in remote areas prevents the deciphering of processes that force the CO_2 flux variability. In this study, by examining atmospheric CO_2 measurements from satellites in the period 2009–2018, we find recurrent end-of-dry-season CO_2 pulses over the Australian continent. These pulses largely control the year-to-year variability of Australia's CO_2 balance. They cause two to three times larger seasonal variations compared with previous top-down inversions and bottom-up estimates. The pulses occur shortly after the onset of rainfall and are driven by enhanced soil respiration preceding photosynthetic uptake in Australia's semiarid regions. The suggested continental-scale relevance of soil-rewetting processes has substantial implications for our understanding and modeling of global climate–carbon cycle feedbacks.

Terrestrial ecosystems drive the seasonal and year-to-year variability of the global carbon dioxide (CO_2) sink (*1*). Previous research has identified semiarid regions as hotspots of global CO_2 balance interannual variability (*2*–*5*) because of the strong sensitivity of photosynthetic carbon uptake to fluctuations in water availability (*6*, *7*). The Australian continent is primarily covered with semiarid ecosystems and experiences large variations in rainfall. These conditions make Australia particularly relevant for the variability in the global carbon cycle (*8*–*13*), contributing up to 60% of yearly anomalies in the global terrestrial CO_2 sink (*2*).

However, attribution of global CO_2 sink variations to certain regions and mechanisms is highly uncertain, which limits our ability to model climate–carbon cycle feedbacks and project future climate (*14*, *15*). Global process-based ecosystem models underestimate observed CO_2 flux variability across semiarid sites owing to the complexity of carbon–water cycle interactions and the diversity of ecosystem responses to water fluctuations (*16*, *17*). The same holds true for machine learning–based models trained on local carbon flux observations (*18*, *19*) given the scarcity of available flux measurements in low-latitude semiarid regions (*20*) and the inability to represent potentially important noninstantaneous carry-over effects (*21*). Atmospheric transport inversions based on in situ measurements of airborne CO_2 also suffer from the scarcity of observations in remote areas, and therefore the inversions cannot reliably attribute CO_2 flux variability to specific regions, despite growing monitoring capacities (*22*, *23*). However, recent satellite observations of atmospheric column CO_2 deliver data where ground-based in situ concentration measurements and carbon flux networks are sparse, and thus, satellite CO_2 data can fill important gaps and provide new constraints on regional-scale patterns and processes (*8*, *24*–*28*).

In this study, using satellite observations of atmospheric CO_2 concentrations from the Greenhouse Gases Observing Satellite (GOSAT) for the period 2009–2018, we identify a net CO_2 pulse to the atmosphere of variable magnitude that occurs over Australia at the end of the dry season in most years. We show that this pattern—which is not evident in traditional atmospheric inversions using in situ measurements only, in the FLUXCOM machine learning–based extrapolations of in situ flux measurements (*18*, *20*), or in most process-based ecosystem models of the TRENDY initiative (*29*)—appears to dominate the seasonal and year-to-year variations of Australia's CO_2 balance for that period. The few process-based TRENDY models that reproduce the CO_2 pulse pattern qualitatively suggest that it is caused by rapid respiratory carbon release with the onset of the rainy season while the increase in photosynthetic carbon uptake lags behind. This observed process is consistent with the phenomenon of respiration pulses after rewetting events discussed in the context of the "Birch effect" (*30*, *31*). Such pulses have been described extensively in local studies of water-limited systems (*32*), but their large-scale relevance has remained unknown.

Atmospheric CO_2 peak over Australia

The GOSAT has been delivering global measurements of the column-average dry-air mole fractions ("concentrations") of atmospheric CO_2 since its launch in 2009 (*33*). After subtracting the secular trend (*34*), the record of GOSAT concentrations for the period 2009–2018 (Fig. 1) reveals a seasonal pattern above Australia with CO_2 drawdown in March, April, and May (MAM) and a CO_2 peak of variable magnitude at the end of the dry season in October, November, and December (OND). These patterns are consistent among two retrievals independently applied to GOSAT [GOSAT/RemoTeC (*35*) and GOSAT/ACOS (*36*); table S1], and they are present in CO_2 concentrations measured by the Orbiting Carbon Observatory-2 [OCO-2 (*37*, *38*), period 2015–2018; table S1) as well as in ground-based data of the Total Carbon Column Observing Network (TCCON) (*39*) (figs. S1 and S2).

In contrast, the atmospheric column CO_2 concentrations simulated by three inverse atmospheric transport models—CarbonTracker CT2019B (*40*), CAMS (*41*), and TM5-4DVAR (*42*)—underestimate the CO_2 drawdown in MAM and lack the CO_2 pulses in OND (Fig. 1). Driven by atmospheric wind data, these transport models deliver concentration fields that are optimally compatible with in situ–measured CO_2 concentrations and the a priori biogenic, oceanic, fire, and fossil CO_2 surface-atmosphere fluxes (*34*). However, given their sparsity in

[1]Institute of Environmental Physics, Heidelberg University, 69120 Heidelberg, Germany. [2]Heidelberg Center for the Environment (HCE), Heidelberg University, 69120 Heidelberg, Germany. [3]Goddard Space Flight Center, NASA, Greenbelt, MD 20771, USA. [4]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740, USA. [5]Max Planck Institute for Biogeochemistry, 07745 Jena, Germany. [6]College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4RJ, Devon, UK. [7]Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, BC V8N 1V9, Canada. [8]Climate Science Centre, CSIRO Oceans and Atmosphere, Canberra, ACT 2601, Australia. [9]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QF, UK. [10]Laboratoire de Météorologie Dynamique, Institut Pierre-Simon Laplace, CNRS-ENS-UPMC-X, Département de Géosciences, Ecole Normale Supérieure, 75005 Paris, France. [11]Université Paris Saclay, CEA-CNRS-UVSQ, LSCE/IPSL, 91191 Gif sur Yvette, France. [12]Department of Atmospheric Sciences, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. [13]Institute of Applied Energy, Tokyo 105-0003, Japan. [14]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO 80305, USA. [15]Max Planck Institute for Meteorology, 20146 Hamburg, Germany. [16]Biospheric Sciences Laboratory, Goddard Space Flight Center, NASA, Greenbelt, MD 20771, USA. [17]CNRM, Université de Toulouse, Météo-France, CNRS, 31057 Toulouse, France. [18]Schiller Institute for Integrated Science and Society, Department of Earth and Environmental Sciences, Boston College, Chestnut Hill, MA 02467, USA. [19]Met Office Hadley Centre for Climate Science and Services, Exeter EX1 3PB, UK. [20]School of Atmospheric Sciences, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Sun Yat-sen University, Zhuhai 519082, China. [21]Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology (NUIST), Nanjing 210044, China. [22]Centre for Atmospheric Chemistry, School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, NSW 2522, Australia. [23]Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, 69120 Heidelberg, Germany.
***Corresponding author. Email: eva-marie.metz@iup.uni-heidelberg.de (E.-M.M.); andre.butz@iup.uni-heidelberg.de (A.B.)**

and around Australia (compare fig. S3 and fig. S4), the in situ measurements provide only marginal constraints on the regional flux balance. Thus, the discrepancy between CO_2 concentrations from GOSAT and traditional in situ–based atmospheric inversions hints at the existence of a carbon release mechanism in Australian ecosystems that has remained undetected by the existing in situ CO_2 monitoring system.

Australian top-down and bottom-up fluxes

To improve upon the surface flux estimates for Australia, we fed the GOSAT CO_2 concentrations into one of the atmospheric inverse models (TM5-4DVAR) together with the in situ CO_2 measurements. We find that the recurring end-of-dry-season CO_2 concentration peaks are indeed attributed to a carbon release pattern originating from land ecosystems, which is not present in the inversions when assimilating in situ CO_2 data alone (Fig. 2A and fig. S5).

Our new estimates of Australia's carbon balance variability based on assimilating GOSAT together with in situ data show a nearly doubled peak-to-peak amplitude of the seasonal cycle (175 ± 40 TgC/month, mean ± standard deviation over the 2009–2018 period, July-to-June peak-to-peak amplitude) compared with the in situ–only inversions (88 ± 13 TgC/month). Moreover, the end-of-dry-season CO_2 pulses found by the GOSAT inversions imply a more than fourfold greater year-to-year variability of the annual CO_2 fluxes (0.207 PgC/year, standard deviation over the 2010–2018 period) than for the in situ–only inversions (0.039 PgC/year) (fig. S6 and table S2). Fluxes obtained by assimilating OCO-2 together with in situ data for the period 2015–2018 show the same end-of-dry-season pulses and agree well with the fluxes of the GOSAT inversion (fig. S7).

To understand the origin of the CO_2 pulses, we compared the top-down inversions with bottom-up estimates from machine learning [FLUXCOM (*18*, *20*)] and 18 process-based dynamic global vegetation models (DGVMs) from the TRENDY (v9) ensemble (*42*). The DGVMs also provide the component fluxes of gross primary productivity (GPP) and terrestrial ecosystem respiration (TER) enabling the attribution to variations in photosynthetic carbon uptake and respiratory carbon release. We further included fire emissions (FIRE) from the Global Fire Emission Database (GFED) as a potential factor for explaining the pattern. To compare with the top-down inversions, we calculated net biome production (NBP = TER + FIRE − GPP) by adding fire emissions from GFED to net ecosystem exchange (NEE = TER − GPP) from FLUXCOM. With this sign convention, positive fluxes correspond to carbon emissions into the atmosphere. For TRENDY, NBP was taken directly from the simulations of the DGVMs. We find that FLUXCOM+GFED–derived NBP lacks the end-of-dry-season CO_2 pulses (Fig. 2A), and its seasonal amplitude (64 ± 16 TgC/month) underestimates the one found by the GOSAT inversions by a factor of three. This could be explained by the sparsity of Australian flux tower data in the training of the FLUXCOM machine learning models (only 4 of 224 sites are located in Australia; see fig. S3), causing extrapolation errors (*18*), and by known weaknesses in representing certain fluctuations in response to water availability (*19*) or "memory" effects due to unaccounted carbon pool dynamics (*43*). Our analysis further suggests that local and transported fire emissions might contribute at the beginning of the

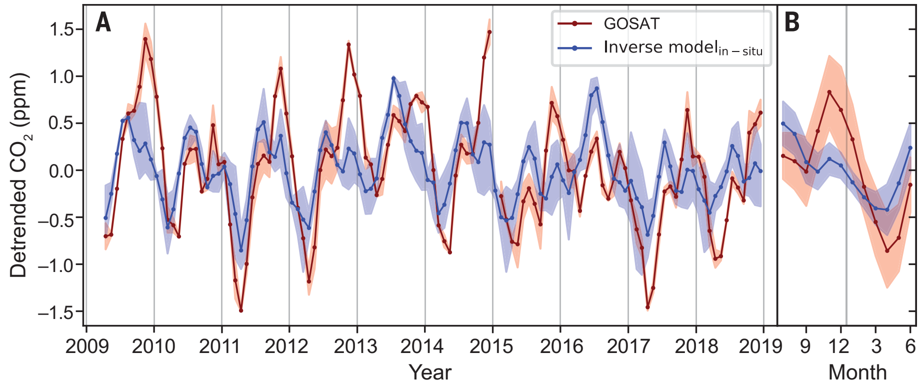


Fig. 1. Detrended CO_2 concentrations over Australia from satellite and models. (**A**) Detrended column-average dry-air mole fractions of CO_2 measured by GOSAT (red) and simulated by inverse models assimilating in situ ground-based measurements (blue). Data are monthly averages for Australia. Red shading indicates the range of the GOSAT/RemoTeC and GOSAT/ACOS algorithms. Blue shading indicates the range of the CarbonTracker, CAMS, and TM5-4DVAR inverse models. ppm, parts per million. (**B**) Mean and standard deviation (shading) over the period 2009–2018.

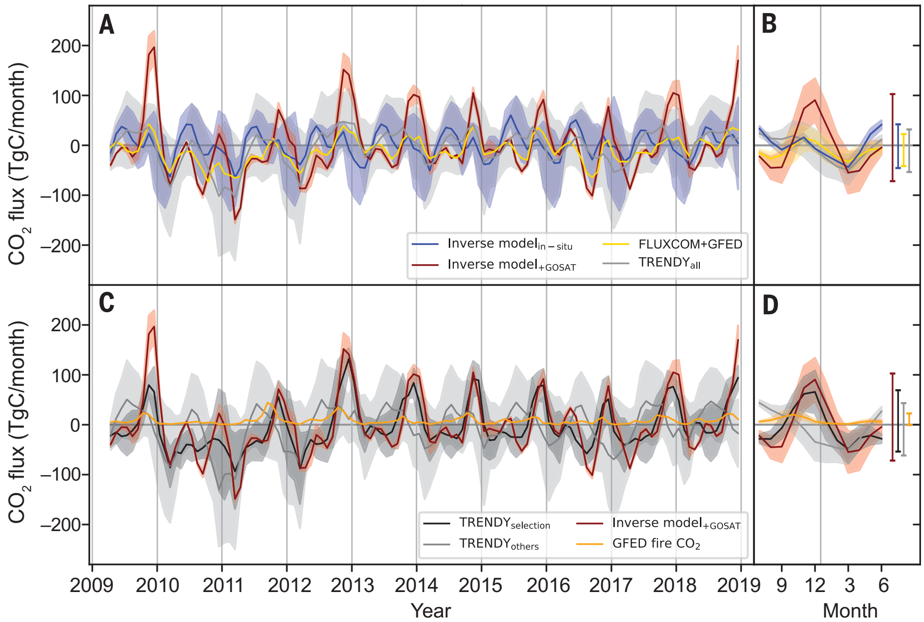


Fig. 2. Australian net CO_2 fluxes. (**A**) Top-down estimates of the net monthly Australian CO_2 fluxes inferred by in situ CO_2 measurement–based inverse models (blue) and by TM5-4DVAR assimilating in situ measurements together with GOSAT observations (red), compared with bottom-up FLUXCOM+GFED NBP (yellow) and the TRENDY ensemble mean NBP (gray). Shading indicates the range among the various top-down data streams (in situ–based CarbonTracker, CAMS, and TM5-4DVAR in blue, TM5-4DVAR$_{+GOSAT/RemoTeC}$ and TM5-4DVAR$_{+GOSAT/ACOS}$ in red) and the standard deviation among the TRENDY ensemble (gray). (**C**) NBP of a subgroup of TRENDY models (black) compared with the other models (gray), with the GOSAT inversions [red, same as in (A)], and with GFED fire emissions (orange). Shading as in (A). (**B** and **D**) Mean and standard deviation (shading) over the period 2009–2018 and the mean peak-to-peak seasonal cycle amplitudes (bars). Positive fluxes correspond to carbon emissions into the atmosphere.

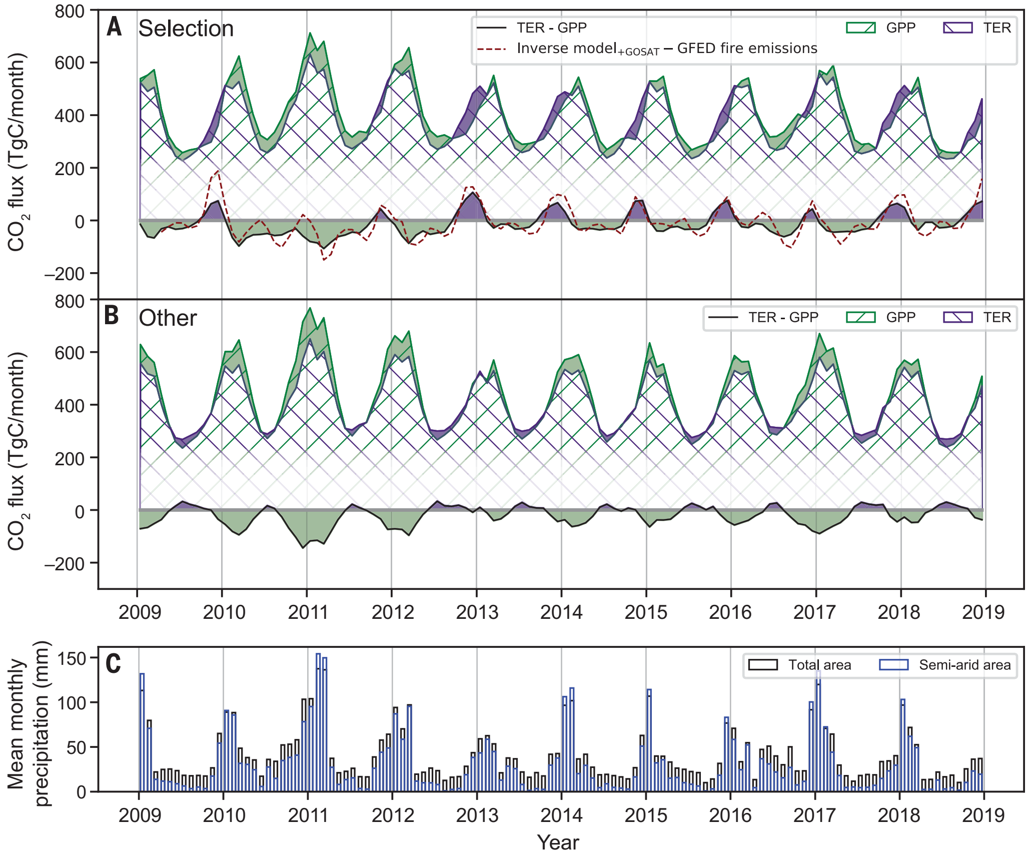


Fig. 3. Seasonal timing of gross CO_2 fluxes among TRENDY models. (**A**) Gross primary production (GPP, green) and total respiration (TER, purple) for Australia for the selection of TRENDY DGVMs that replicate the end-of-dry-season CO_2 pulses. The difference of TER and GPP is given in black in the lower part together with GOSAT-based inversion where GFED fire emissions are subtracted (dashed red line). (**B**) Same as (A), but for the other TRENDY models that do not replicate the end-of-dry-season CO_2 pulses. (**C**) Mean monthly precipitation over the entire Australian region (black) and the semiarid part (fig. S3) of Australia (blue).

carbon pulses but cannot explain their magnitude and duration (Fig. 2B and fig. S8).

The ensemble of TRENDY NBP simulations shows a large intermodel spread and also no end-of-dry-season CO_2 pulses on average (Fig. 2A) causing a seasonal amplitude (85 ± 20 TgC/month) that is about half of that of the GOSAT inversions. However, the dry season pulses are present in a subset of five of the TRENDY DGVMs [Fig. 2B and table S1; "Characteristics of $TRENDY_{selection}$" in (*34*)]. For this subset, the timing, duration, and magnitude (except for the year 2009) of the pulses and their seasonal amplitude (123 ± 31 TgC/month) are closer to the pulses found by the GOSAT inversions. This finding suggests that the CO_2 pulses can be explained by ecosystem processes shaping the phasing of photosynthesis and respiration.

Phasing of respiration and photosynthesis

We find that the subset of DGVMs that are in good agreement with the GOSAT inversions reveal a distinctly different seasonal timing of GPP and TER than the other DGVMs. For the selected subset, the CO_2 pulses are driven by TER, which increases rapidly at the onset of the rainy season, whereas GPP picks up only a few weeks later (Fig. 3A). The pulses originate mainly from an early increase of soil respiration in semiarid regions (figs. S9 and S10A). For the other DGMVs, TER and GPP show a mostly synchronous phasing throughout the year, yielding no CO_2 pulses (Fig. 3B and fig. S10B). The precipitation records for the semiarid regions of Australia (Fig. 3C and fig. S3) suggest that the soil respiration–driven pulses shown by the GOSAT inversions and the selected TRENDY models are weaker or do not occur in years with anomalously strong precipitation during the dry period (Austral winter), such as in the La Niña years 2010 and 2016. This implies that the observed pulses are conditional on rewetting of dry soils and that it is through the strength of the pulses that climatic conditions exert control over Australia's annual CO_2 balance (fig. S6).

The detected continental-scale CO_2 pulses are consistent with site-level observations of dryland ecosystems that show an asynchronous response of respiration and photosynthesis to precipitation pulses (*44*). The rapid response of microbial respiration to rewetting events is known as the "Birch effect" and has been described in the literature of specific sites in some semiarid regions for many decades (*30–32*). After being dormant in the dry period, soil microbes are activated by the moisture supply from rainfall. Under warm soil temperatures, the microbes quickly respire accumulated and readily available substrate and their populations grow rapidly. These dynamics of soil microbial processes cause respiration CO_2 pulses with rewetting of dry soils that are evident in Australian flux tower data (figs. S11 and S12). Photodegradation of surface litter (*45*) and the death of microorganisms during the dry period (*46*, *47*) may lead to the accumulation of easily decomposable substrate available to microorganisms at the onset of rain. It remains an open question whether the respiration pulses are mainly driven by substrates accumulated during the dry period and to what extent they are fueled by mobilization and decomposition of physically protected carbon (*47*). These processes are not represented explicitly or in detail in the TRENDY DGVMs, and thus the DGVMs cannot resolve how the site-level mechanisms scale up to the continental-scale effect observed here. Nonetheless, a selection of models effectively captures the continental-scale CO_2 pulses by a fast response of respiration and a delayed response of photosynthesis to the onset of the rainy season. This highlights the importance of subtle differences in effective parameterizations of respiration and photosynthesis to moisture fluctuations. Associated uncertainties affect the ability of the models to accurately represent the carbon cycle of semiarid ecosystems.

Our study demonstrates that the soil respiration–driven CO_2 pulses over Australia after the end of the dry season are of large-scale relevance and appear to dominate the variability of the continent's carbon balance.

The GOSAT inversions have shed light on a blind spot of previous top-down and bottom-up approaches for quantifying and attributing CO_2 flux variability. This is important because Australia's semiarid regions contribute substantially to the interannual variability of the global terrestrial carbon sink and because it is the ecosystem response to the phasing of dry and wet periods that drives the seasonal mechanism behind the large interannual variability. Thus, our study calls for revisiting the contributions of global semiarid systems to CO_2 balance variations and for assessing implications for our ability to model climate-carbon feedbacks in semiarid regions. Only a few of the global vegetation models are able to reproduce the observed CO_2 pulses, which suggests that only their respective parameterizations are able to represent the sensitivity of the underlying mechanism to changes in climatic conditions and, thus, to accurately project semiarid CO_2 flux variability under a changing climate. Considering the large uncertainties associated with modeling climate-carbon feedbacks (*14*, *15*, *48*), our findings may contribute continental-scale mechanistic understanding that can help reduce these uncertainties for dryland ecosystems that are found to be particularly sensitive to climate change (*49*).

REFERENCES AND NOTES

1. P. Friedlingstein *et al*., *Earth Syst. Sci. Data* **12**, 3269–3340 (2020).
2. B. Poulter *et al*., *Nature* **509**, 600–603 (2014).
3. A. Ahlström *et al*., *Science* **348**, 895–899 (2015).
4. M. Jung *et al*., *Nature* **541**, 516–520 (2017).
5. V. Humphrey *et al*., *Nature* **592**, 65–69 (2021).
6. S. Piao *et al*., *Glob. Change Biol.* **26**, 300–318 (2020).
7. V. Haverd, A. Ahlström, B. Smith, J. G. Canadell, *Glob. Change Biol.* **23**, 793–800 (2017).
8. R. G. Detmers *et al*., *Geophys. Res. Lett.* **42**, 8177–8184 (2015).
9. X. Ma *et al*., *Sci. Rep.* **6**, 37747 (2016).
10. J. Cleverly *et al*., *Sci. Rep.* **6**, 23113 (2016).
11. Z. Xie *et al*., *Remote Sens. Environ.* **231**, 111270 (2019).
12. A. Bastos *et al*., *Global Biogeochem. Cycles* **34**, e2019GB006393 (2020).
13. L. Teckentrup *et al*., *Biogeosciences* **18**, 5639–5668 (2021).
14. P. M. Cox, C. Huntingford, M. S. Williamson, *Nature* **553**, 319–322 (2018).
15. S. Wenzel, P. M. Cox, V. Eyring, P. Friedlingstein, *Nature* **538**, 499–501 (2016).
16. D. L. Hoover, A. A. Pfennigwerth, M. C. Duniway, *J. Ecol.* **109**, 3280–3294 (2021).
17. N. MacBean *et al*., *Environ. Res. Lett.* **16**, 094023 (2021).
18. G. Tramontana *et al*., *Biogeosciences* **13**, 4291–4313 (2016).
19. P. Bodesheim, M. Jung, F. Gans, M. D. Mahecha, M. Reichstein, *Earth Syst. Sci. Data* **10**, 1327–1365 (2018).
20. M. Jung *et al*., *Biogeosciences* **17**, 1343–1365 (2020).
21. S. Sippel *et al*., *Curr. Clim. Change Rep.* **4**, 266–286 (2018).
22. J. Beringer *et al*., *Biogeosciences* **13**, 5895–5916 (2016).
23. J. Cleverly *et al*., *Environ. Res. Lett.* **14**, 095004 (2019).
24. P. J. Sellers, D. S. Schimel, B. Moore3rd, J. Liu, A. Eldering, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7860–7868 (2018).
25. P. I. Palmer *et al*., *Nat. Commun.* **10**, 3344 (2019).
26. B. Byrne *et al*., *J. Geophys. Res. Atmos.* **125**, e2019JD032029 (2020).
27. Z. Chen *et al*., *Environ. Res. Lett.* **16**, 054041 (2021).
28. Y. Villalobos, P. Rayner, S. Thomas, J. Silver, *Atmos. Chem. Phys.* **20**, 8473–8500 (2020).
29. S. Sitch *et al*., *Biogeosciences* **12**, 653–679 (2015).
30. H. F. Birch, *Plant Soil* **20**, 43–49 (1964).
31. P. Jarvis *et al*., *Tree Physiol.* **27**, 929–940 (2007).
32. P. Casals, L. Lopez-Sangil, A. Carrara, C. Gimeno, S. Nogués, *Global Biogeochem. Cycles* **25**, GB3012 (2011).
33. A. Kuze *et al*., *Atmos. Meas. Tech.* **9**, 2445–2461 (2016).
34. Materials and methods are available as supplementary materials.
35. A. Butz *et al*., *Geophys. Res. Lett.* **38**, L14812 (2011).
36. T. E. Taylor *et al*., *Earth Syst. Sci. Data* **14**, 325–360 (2022).
37. A. Eldering *et al*., *Science* **358**, eaam5745 (2017).
38. Y. Villalobos *et al*., *Atmos. Chem. Phys. Discuss.* **22**, 8897–8934 (2022).
39. D. Wunch *et al*., *Philos. Trans. R. Soc. London Ser. A* **369**, 2087–2112 (2011).
40. W. Peters *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18925–18930 (2007).
41. F. Chevallier *et al*., *J. Geophys. Res.* **115**, D21307 (2010).
42. S. Basu *et al*., *Atmos. Chem. Phys.* **13**, 8695–8717 (2013).
43. S. Besnard *et al*., *PLOS ONE* **14**, e0211510 (2019).
44. T. E. Huxman *et al*., *Oecologia* **141**, 254–268 (2004).
45. F. E. Moyano, S. Manzoni, C. Chenu, *Soil Biol. Biochem.* **59**, 72–85 (2013).
46. W. Borken, E. Matzner, *Glob. Change Biol.* **15**, 808–824 (2009).
47. J. P. Schimel, *Annu. Rev. Ecol. Evol. Syst.* **49**, 409–432 (2018).
48. V. K. Arora *et al*., *Biogeosciences* **17**, 4173–4222 (2020).
49. X. Lian *et al*., *Nat. Rev. Earth Environ.* **2**, 232–250 (2021).
50. A. Butz, RemoTeC full-physics retrieval GOSAT/TANSO-FTS Level 2 bias-corrected XCO2 version 2.4.0 operated at Heidelberg University, RemoTeCv2.4.0, Zenodo (2022); https://doi.org/10.5281/zenodo.5886662.
51. E.-M. Metz, ATMO-IUP-UHEI/MetzEtAl2023: v1.0.0, version 1.0.0, Zenodo (2023); https://doi.org/10.5281/zenodo.7648699.

ACKNOWLEDGMENTS

We gratefully acknowledge the data storage service SDS@hd supported by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) and the computing resources provided by the DKRZ under project bb1170. E.-M.M. acknowledges a doctoral scholarship from the German National Academic Foundation. We thank the Japanese Aerospace Exploration Agency, the National Institute for Environmental Studies, and the Ministry of Environment for the GOSAT data and their continuous support as part of the Joint Research Agreement. We thank the OCO-2 science team for producing the GOSAT/ACOS L2 XCO2 data. OCO-2 data were produced by the OCO-2 project at the Jet Propulsion Laboratory, California Institute of Technology, and obtained from the OCO-2 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center. CarbonTracker CT2019B results were provided by NOAA ESRL, Boulder, Colorado, USA, from the website at http://carbontracker.noaa.gov. We thank all TRENDY modelers for providing model output as part of the TRENDY v9 ensemble. R.S. thanks the Météo-France/DSI supercomputing center, which provided computing time for performing TRENDY simulations. D.S.G. acknowledges support from the ANR CLAND Convergence Institute. This study has greatly benefited from discussions with C. Frankenberg. **Funding:** Funding was provided by the German Research Foundation (DFG) through grants BU2599/1-1 and INST 35/1503-1 FUGG and by the Darwin and Wollongong TCCON stations through ARC grants DP160100598, LE0668470, DP140101552, DP110103118, and DP0879468 and Darwin through NASA grants NAG5-12247 and NNG05-GD07G. S.Z. was supported by EU H2020 grant 821003, 4C; S.B. was supported by NASA grant 80NSSC20K0818; and R.S. was supported by EU H2020 CRESCENDO (641816) and ESM2025 (101003536). **Author contributions:** A.B., S.N.V., and E.-M.M. were involved in conceptualization and methodology. E.-M.M. conducted the formal analysis and the visualization under the supervision of A.B. and S.N.V. A.B., S.N.V., E.-M.M., M.J., and S.B. wrote the original draft. S.B. performed the dedicated TM5-4DVAR runs. S.S., V.K.A., P.R.B., P.F., D.S.G., A.K.J., E.K., D.L., J.E.M.S.N., B.P., R.S., H.T., A.W., W.Y., X.Y., and S.Z. provided TRENDY data. N.M.D. and D.W.T.G. provided TCCON data. All authors contributed to the editing and review of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** GOSAT/RemoTeC2.4.0 XCO2 data can be obtained from Zenodo (*50*) (last access: 2022-02-25). GOSAT/ACOS data are available at https://oco2.gesdisc.eosdis.nasa.gov/data/GOSAT_TANSO_Level2/ACOS_L2_Lite_FP.9r/ (last access: 2020-07-28). OCO-2 data are available at https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_10r/summary (last access: 2020-11-01). TCCON data can be downloaded at https://data.caltech.edu/records/269 (last access: 2022-02-25). CarbonTracker CT2019B CO2 fluxes and concentrations can be downloaded from https://gml.noaa.gov/aftp/products/carbontracker/co2/CT2019B/fluxes/monthly/ (last access: 2021-02-19) and https://gml.noaa.gov/aftp/products/carbontracker/co2/CT2019B/molefractions/co2_total_monthly/ (last access: 2022-02-25), respectively. CAMS concentrations and fluxes can be found at https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-greenhouse-gas-inversion (last access: 2021-10-07). GFAS emissions records are available at https://apps.ecmwf.int/datasets/data/cams-gfas/ (last access: 2020-11-13). CAMS and GFAS data were generated using Copernicus Atmosphere Service Information [2021], and neither the European Commission nor the European Centre for Medium-Range Weather Forecasts (ECMWF) is responsible for any use that may be made of the information it contains. GFED fire emissions are available at https://www.geo.vu.nl/~gwerf/GFED/GFED4/ (last access: 2020-07-10). FINN data were retrieved from the American National Center for Atmospheric Research https://www2.acom.ucar.edu/modeling/finn-fire-inventory-ncar (last access: 2020-11-18). The MIP data used in this work can be downloaded from https://www.gml.noaa.gov/ccgg/OCO2_v10mip/ (last access: 2022-11-06). ERA5-land data records contain modified Copernicus Atmosphere Service Information [2021] available at the Climate Data Store https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means (last access: 2021-12-20). This work used eddy covariance data collected by the TERN-OzFlux facility. OzFlux acknowledges the financial support of the Australian Federal Government via the National Collaborative Research Infrastructure Scheme and the Education Investment Fund. OzFlux data are available at https://data.ozflux.org.au (last access: 2023-01-20). TRENDYv9 model output and FLUXCOM products are available at https://sites.exeter.ac.uk/trendy and http://fluxcom.org/CF-Download/, respectively. Data S1 to S3 in the supplementary materials contain monthly TRENDY and FLUXCOM data used in this study. Monthly TM5-4DVAR data are available in data S1 and S4. The code used in this study is available at Zenodo (*51*) or GitHub (https://github.com/ATMO-IUP-UHEI/MetzEtAl2023).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add7833
Materials and Methods
Figs. S1 to S12
Tables S1 and S2
References (*52–132*)
Data S1 to S4

Submitted 4 July 2022; accepted 6 March 2023
10.1126/science.add7833

Check for updates

TAU IMMUNOTHERAPY

Cytosolic antibody receptor TRIM21 is required for effective tau immunotherapy in mouse models

Aamir S. Mukadam[1,5]*, Lauren V. C. Miller[1,5]†, Annabel E. Smith[1,5]†, Marina Vaysburd[2], Siri A. Sakya[3,4], Sophie Sanford[1,5], Sophie Keeling[1,5], Benjamin J. Tuck[1,5], Taxiarchis Katsinelos[1,5], Chris Green[1,5], Lise Skov[1,5], Sanne S. Kaalund[5]‡, Stian Foss[3,4], Keith Mayes[2], Kevin O'Connell[2], Mark Wing[2], Claire Knox[2], Jessica Banbury[2], Edward Avezov[1,5], James B. Rowe[5,6], Michel Goedert[2], Jan Terje Andersen[3,4], Leo C. James[2], William A. McEwan[1,5]*

Aggregates of the protein tau are proposed to drive pathogenesis in neurodegenerative diseases. Tau can be targeted by using passively transferred antibodies (Abs), but the mechanisms of Ab protection are incompletely understood. In this work, we used a variety of cell and animal model systems and showed that the cytosolic Ab receptor and E3 ligase TRIM21 (T21) could play a role in Ab protection against tau pathology. Tau-Ab complexes were internalized to the cytosol of neurons, which enabled T21 engagement and protection against seeded aggregation. Ab-mediated protection against tau pathology was lost in mice that lacked T21. Thus, the cytosolic compartment provides a site of immunotherapeutic protection, which may help in the design of Ab-based therapies in neurodegenerative disease.

Several neurodegenerative diseases are characterized by the progressive accumulation of cytosolic assemblies of hyperphosphorylated tau (*1*). Extracellular tau assemblies are taken up into recipient cells after interactions with surface heparan sulfate proteoglycans and low-density lipoprotein receptor-related protein 1 (LRP1) (*2*–*5*) and can induce the seeded aggregation of native tau pools (*6*). The passive transfer of antibodies (Abs) against tau can reduce tau pathology in animal models and is under investigation as a disease-modifying treatment in humans (*7*–*11*). The mechanisms of this protection remain uncertain, with roles for microglial internalization that uses cell-surface Ab receptors, Fc-γ receptors (FcγRs) (*12*–*14*), blocking of seed entry to cells (*13*, *15*, *16*), and endolysosomal degradation (*17*–*19*) suggested as potential modes (*20*). The cytosolic Ab receptor and E3 ubiquitin ligase, TRIM21 (T21), engages Ab-bound particles inside the cell and mounts a potent degradation response at the proteasome (*21*–*23*). In cell-based assays, the introduction of Abs to cells can induce the T21-dependent selective degradation of numerous cellular proteins, including tau (*24*–*26*). However, the contribution of T21 to immunotherapeutic protection against tau pathology remains undetermined.

[1]UK Dementia Research Institute at the University of Cambridge, Cambridge CB2 0AH, UK. [2]MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK. [3]Department of Immunology, University of Oslo and Oslo University Hospital Rikshospitalet, N-0424 Oslo, Norway. [4]Institute of Clinical Medicine and Department of Pharmacology, University of Oslo and Oslo University Hospital, N-0372 Oslo, Norway. [5]Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 0AH, UK. [6]Cambridge University Hospitals NHS Trust, Cambridge CB2 0SZ, UK.
*Corresponding author. Email: asm64@cam.ac.uk (A.S.M.); wm305@cam.ac.uk (W.A.M.)
†These authors contributed equally to this work. ‡Present address: Research Laboratory for Stereology and Neuroscience, Bispebjerg-Frederiksberg Hospital, Copenhagen, Denmark.

Tau assemblies enter neurons in complex with Abs to contact T21

To investigate whether seeded aggregation of tau may be intercepted by T21, we first assessed whether Abs could be taken into neurons in complex with tau assemblies. We incubated recombinant heparin-assembled tau with BR134, a rabbit polyclonal Ab raised against the C terminus of tau (*27*) that can neutralize seeded tau aggregation in human cell lines (*25*). After 8 hours, tau assemblies were observed within neurons irrespective of whether Ab was present, which indicates that BR134 did not prevent their uptake (Fig. 1, A and B). When BR134 was present, these intracellular tau assemblies colocalized with T21, which resides in the cytosol. The number of intracellular T21-positive tau assemblies increased over the course of 8 hours, which is consistent with the entry dynamics of tau to the cytosol of neurons (Fig. 1C and fig. S2) (*28*). Dimers of T21 bind Abs through interactions between the T21 PRYSPRY domain and the Ab Fc region (*29*). We confirmed this interaction in the context of BR134 and the mouse T21 PRYSPRY domain using fluorescence anisotropy and observed a monomer dissociation constant (K_d) of 19 nM (Fig. 1D). Thus, tau assemblies can enter the cytosol of neurons in complex with antibodies and recruit T21 through a high-affinity interaction between the Ab Fc region and the T21 PRYSPRY domain.

T21 and Abs lead to functional inactivation of tau seeding behavior

We next investigated whether T21 contributes to the neutralization of seeded tau aggregation. We used an organotypic hippocampal slice culture (OHSC) model of seeding (*30*). Slices were prepared from transgenic mice that expressed human tau with frontotemporal lobar degeneration–associated mutation P301S (P301S Tau-Tg) and cultured at the air-liquid interface (Fig. 1E). We also prepared slices from a T21-deficient mouse line (*31*) on the same tau background (P301S Tau-Tg T21$^{-/-}$) (*32*). OHSCs derived from both genotypes retained normal representation of the major cell types of the central nervous system, and OHSCs from T21$^{-/-}$ animals did not express T21 detectable with immunoblot (fig. S3). OHSCs maintained tau in a native state over 8 weeks in culture, and the genotypes displayed similar levels of seeded aggregation in response to heparin-assembled tau (fig. S3). However, there was a substantial difference in the observed levels of neutralization by BR134 between the genotypes (Fig. 1, F and G). BR134 reduced seeding by >90% in T21$^{+/+}$ OHSCs compared with control Abs. However, genetic deletion of T21 almost completely abolished the ability of BR134 to prevent seeded aggregation. We next examined whether the activity of Abs and T21 could inhibit the formation of seed-competent tau species that occurs as a result of seeded aggregation in the OHSCs. We treated OHSCs with recombinant heparin tau assemblies in the presence and absence of BR134 as above. OHSC homogenates were examined 3 weeks later for the levels of seed-competent species on a sensitive reporter cell line [human embryonic kidney (HEK) 293 P301S tau-venus (*25*)]. Untreated OHSCs contained only low levels of tau seeds, whereas those treated with tau assemblies induced substantial levels of seeded aggregation (Fig. 1H). We observed a significant reduction in tau seeds in response to treatment with BR134 when compared with control Abs, but only when T21 was present. In P301S Tau-Tg T21$^{-/-}$ OHSCs, Abs were unable to reduce the number of seeds that were produced. T21 promotes virus neutralization through its E3 ligase activity, which stimulates degradation with the involvement of the ubiquitin-proteasome system (*21*). To test whether this activity was required for the neutralization of seeding, we used TAK-243, an inhibitor of UBA1, an E1 ubiquitin-activating enzyme (*33*), which prevented polyubiquitination in primary neurons (fig. S4). Neutralization of tau seeding was no longer observed when the inhibitor was applied (Fig. 1I and fig. S4). Thus, Abs recruit T21 to internalized tau assemblies, inhibit the formation of seed-competent tau assemblies, and reduce the levels of hyperphosphorylated tau inclusions.

Comparison of T21 with classical Fc receptors in organotypic slice culture

Abs can mediate extracellular protection against tau by promoting uptake to microglia through interactions with FcγRs (*12*, *14*). We thus examined the contribution of FcγR interaction in

 https://avxhm.se/blogs/hill0

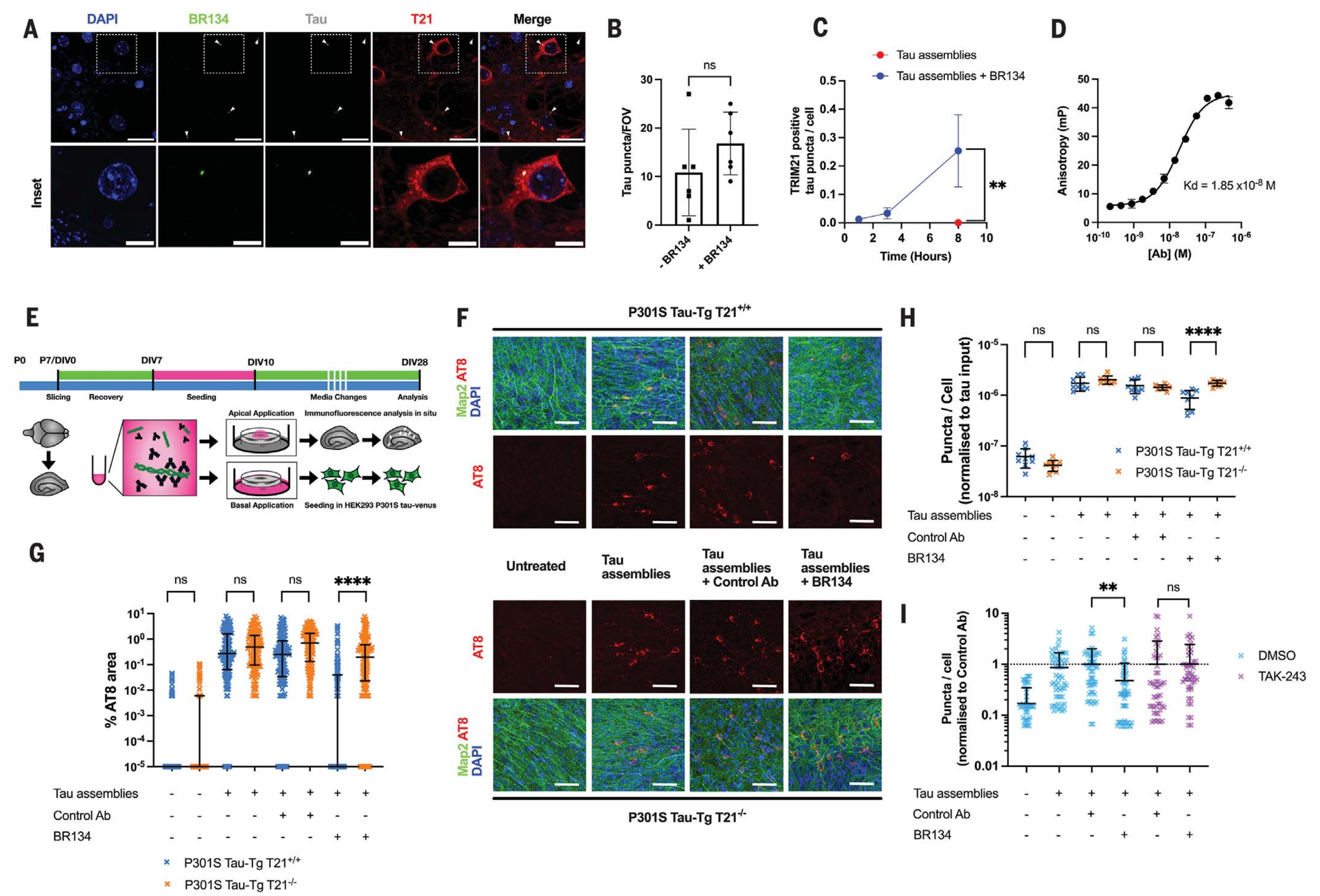


Fig. 1. Mechanisms of Ab protection in neuronal cultures. (**A**) Confocal immunofluorescence microscope images of mouse primary neurons expressing mCherry-T21 treated with tau assemblies in complex with tau C terminus–specific rabbit polyclonal Ab, BR134. Arrows indicate intracellular Ab-tau assembly complexes, most of which were found to colocalize with T21. Scale bar, 25 μm; inset scale bar, 10 μm. (**B**) Number of tau assemblies detectable within neurons 8 hours after their addition in the presence or absence of BR134. (**C**) Number of intracellular tau puncta that colocalized with T21 in the presence or absence of BR134. (**D**) Fluorescence anisotropy of Alexa488-labeled mouse T21 PRYSPRY domain in the presence of indicated concentration of BR134. (**E**) Diagram of OHSC model of seeded tau aggregation. Tau assemblies are pretreated with Abs and provided to hippocampal slices prepared from P301S Tau-Tg animals on day in vitro (DIV) 7. OHSCs are fixed for immunofluorescence analysis of tau pathology (AT8) on DIV28 or lysed and examined for levels of tau seeding in HEK293 P301S tau-venus reporter cells. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; G, Gly; H, His; L, Leu; P, Pro; and S, Ser. In the mutants, other amino acids were substituted at certain locations; for example, P301S indicates that proline at position 301 was replaced by serine. (**F**) Representative immunofluorescence images for AT8-reactive tau structures in OHSCs from P301S Tau-Tg $T21^{+/+}$ and $T21^{-/-}$ backgrounds challenged with tau assemblies that were untreated or incubated with control Ab 9C12 or BR134. Map2 staining reveals neuronal architecture. Scale bar, 50 μm. DAPI, 4′,6-diamidino-2-phenylindole. (**G**) Levels of AT8 reactivity in OHSCs after treatment with tau assemblies in the absence of Abs or after preincubation with the indicated Ab. ns, not significant. (**H**) Levels of tau seeds present within OHSCs three weeks after treatment with indicated tau and Ab complexes. Levels were determined by applying OHSC homogenates to HEK293 cells that expressed P301S tau-venus. (**I**) Levels of AT8-reactive tau structures in primary neurons that were derived from P301S Tau-Tg mice challenged with tau assemblies that were untreated or incubated with the indicated Ab in the presence of dimethyl sulfoxide (DMSO) or E1 inhibitor TAK-243. Data normalized to control antibody. Median and interquartile range indicated. (B) and (C) Mean ± SD from $n = 6$ randomly selected fields of view; (D) mean ± SD from $n = 2$ repeats. (G) Points indicate 100- by 100-μm sections from images of OHSCs prepared from $n = 6$ mice with median ± interquartile range. (H) Each point indicates seeding from pooled OHSC homogenates derived from $n = 3$ mice with median ± interquartile range. (I) Points indicate values from individual fields of view from $n = 3$ independent repeats with mean ± SD. (B) and (C) Mann-Whitney test; (G) and (I) Kruskal-Wallis test with Dunn's correction for multiple comparisons; (H) one-way analysis of variance (ANOVA) with Tukey's correction for multiple comparisons. $^{**}P < 0.01$; $^{****}P < 0.0001$.

preventing seeded aggregation in OHSCs. We used a mouse monoclonal Ab, anti-pS422 tau (AP422), which binds to tau phosphorylated at S422 (*34*) and detects tau prepared from Alzheimer's disease, corticobasal degeneration, and progressive supranuclear palsy brains (fig. S5). We used kinase ERK2 to phosphorylate recombinant tau, which generated the AP422 epitope (fig. S6). As with BR134, AP422 protected against seeding and generation of seed-competent species in OHSCs, both of which were dependent on T21 (Fig. 2, A and B, and fig. S5). To enable reverse genetic mutagenesis of Ab Fc region, AP422 was cloned and expressed as recombinant mouse immunoglobulin G2a (IgG2a) (rAP422). We verified that specificity for phospho-tau was maintained (fig. S6) and introduced the Fc amino acid substitutions P329G, L234A, and L235A (PGLALA), which abrogate FcγR interactions (*35*, *36*). We confirmed with enzyme-linked immunosorbent assay (ELISA) that the PGLALA substitutions ablated interactions with mouse FcγRI, FcγRIIB, FcγRIII, and FcγRIV but maintained T21 and

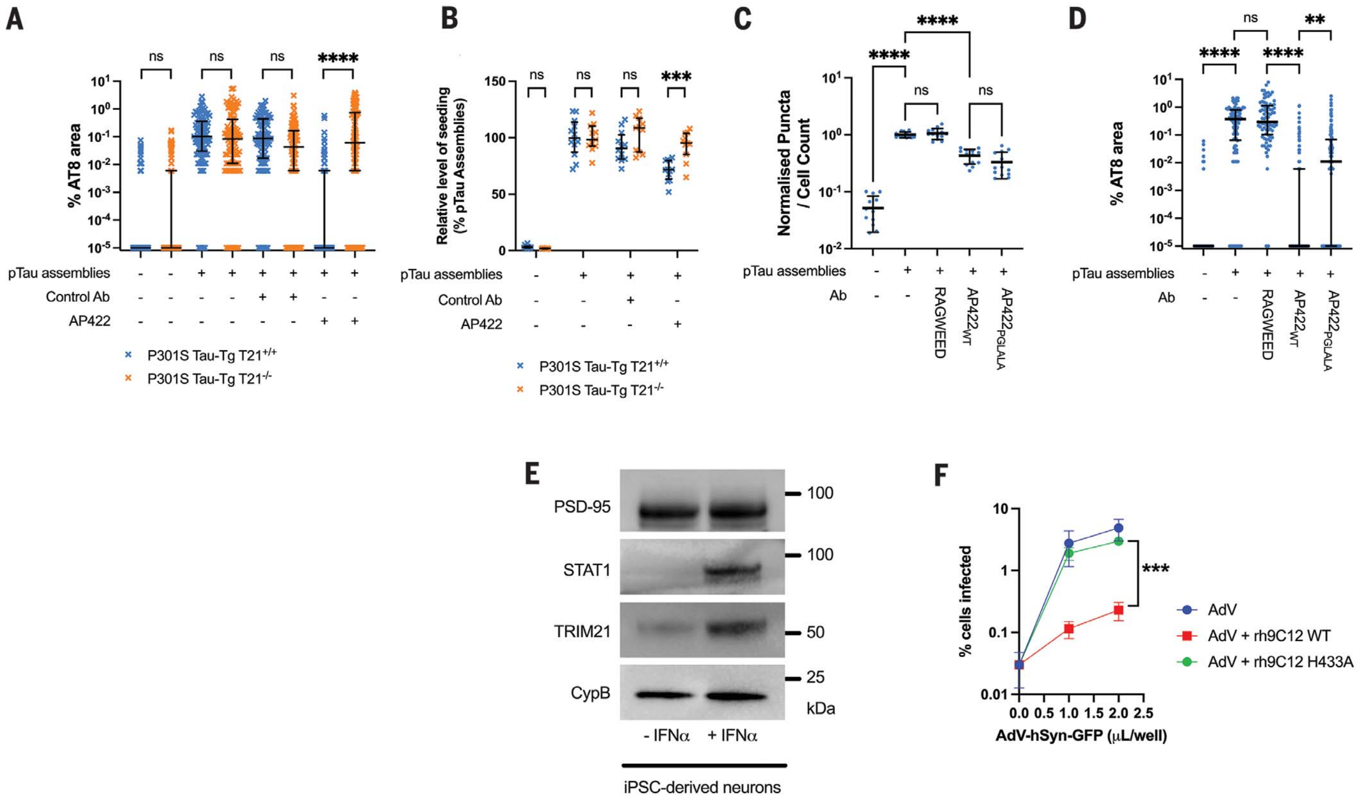


Fig. 2. FcγR contributions in OHSCs and T21 in human iPSC-derived neurons. (**A**) Levels of AT8 staining in P301S Tau-Tg T21$^{+/+}$ and T21$^{-/-}$ OHSCs treated with phospho-tau assemblies in the presence of AP422, a mouse IgG1 that binds to tau phosphorylated at S422, or isotype-matched anti-adenovirus control, 9C12. (**B**) Levels of seeding observed in extracts prepared from OHSCs treated with the indicated tau assemblies and Abs. (**C**) Levels of seeded aggregation in HEK293 cells treated with 1 nM phospho-tau assemblies in the presence of indicated Abs. (**D**) Levels of AT8 staining in OHSCs treated with phospho-tau assemblies that were incubated with the indicated recombinant Abs. Ragweed, anti-ragweed pollen control; AP422$_{WT}$, mouse IgG2a; AP422$_{PGLALA}$, mouse IgG2a with the PGLALA mutations that prevent FcγR interaction. (**E**) Immunoblots for T21, synaptic marker PSD-95, IFN-stimulated gene STAT-1, and loading control CypB in human iPSC-derived neurons in the presence and absence of IFNα. (**F**) Levels of adenovirus type 5 infection in human iPSC-derived neurons in the absence of Ab or in the presence of recombinant anti-AdV 9C12 expressed with human IgG1 Fc. Wild-type Fc or Fc bearing H433A that prevents interaction with TRIM21 was used. GFP, green fluorescent protein. (A) to (D) Median and interquartile range shown. In (A) and (D) points indicate 100- by 100-μm sections from images of OHSCs prepared from n = 6 mice. (B) Each point indicates seeding from pooled OHSC homogenates derived from n = 3 mice. (C) n = 3 biological replicates; points indicate technical replicates. (A) and (D) Kruskal-Wallis test with Dunn's correction for multiple comparisons. (B) and (C) One-way ANOVA with Tukey's correction for multiple comparisons. (F) Mean and SD; n = 3 independent replicates; unpaired t test. $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

FcRn interactions (fig. S6). As a control, we used recombinant mouse IgG2a against ragweed pollen. As expected, in HEK293 cells, which do not express significant levels of FcγRs (*25*), there was no difference between neutralization with rAP422-wild type (WT) versus rAP422-PGLALA (Fig. 2C). In OHSCs, where microglia are present, rAP422-PGLALA was able to exert similar levels of protection as unmodified rAP422-WT, with only a minor portion of its activity being lost (Fig. 2D and fig. S6). This contrasts with T21 deletion, for which neutralization was almost entirely lost. Thus, intracellular neutralization by means of T21 is primarily responsible for protection against tau seeding by AP422 in ex vivo brain slice cultures.

T21 is present and functional in human iPSC-derived neurons

An important consideration for tau immunotherapy in neurodegenerative disease is the level and activity of T21 in human neurons, the major site of tau expression and aggregation in Alzheimer's disease and many other tauopathies. We used human induced pluripotent stem cell (iPSC)–derived neurons to examine whether T21 is available for Ab-mediated degradation in this setting. Immunoblot confirmed that T21 is expressed in human iPSC-derived neurons and is up-regulated by treatment with interferon-α (IFNα), a cytokine known to regulate T21 expression (Fig. 2E) (*21*, *37*). We used neutralization of an adenovirus type 5 vector that expressed green fluorescent protein under the control of a neuron-specific synapsin promoter (AdV) to determine T21 activity. Treatment of AdV with the anti-hexon mouse monoclonal Ab 9C12 neutralizes infection in a T21-dependent manner that can be reversed by the Fc amino acid substitution H433A at the T21 binding interface (*22*). Using a chimeric mouse-human variant of 9C12 with human Fc region (rh9C12) (*38*), we observed potent neutralization of infection of AdV infection in human neurons (Fig. 2F). However, rh9C12 with point mutation H433A was almost completely unable to neutralize infection. Thus, T21 is expressed and active in human neurons, and the T21 pathway is available for engagement by immunotherapy in this cell type.

Contribution of T21 to in vivo protection during tau immunotherapy

We next investigated the role of T21 in a transgenic animal model of tau pathology. In P301S-Tg mice, incipient tau pathology can be detected by immunoreactivity to phospho-tau in lumbar spinal sections from 1 month followed by amplification of signal until 7 months (*39*). Sarkosyl-insoluble tau and seed-competent species increased between 20 and 80 days of age (Fig. 3, A to C). No seed-competent species were detected in nontransgenic C57BL/6

 https://avxhm.se/blogs/hill0

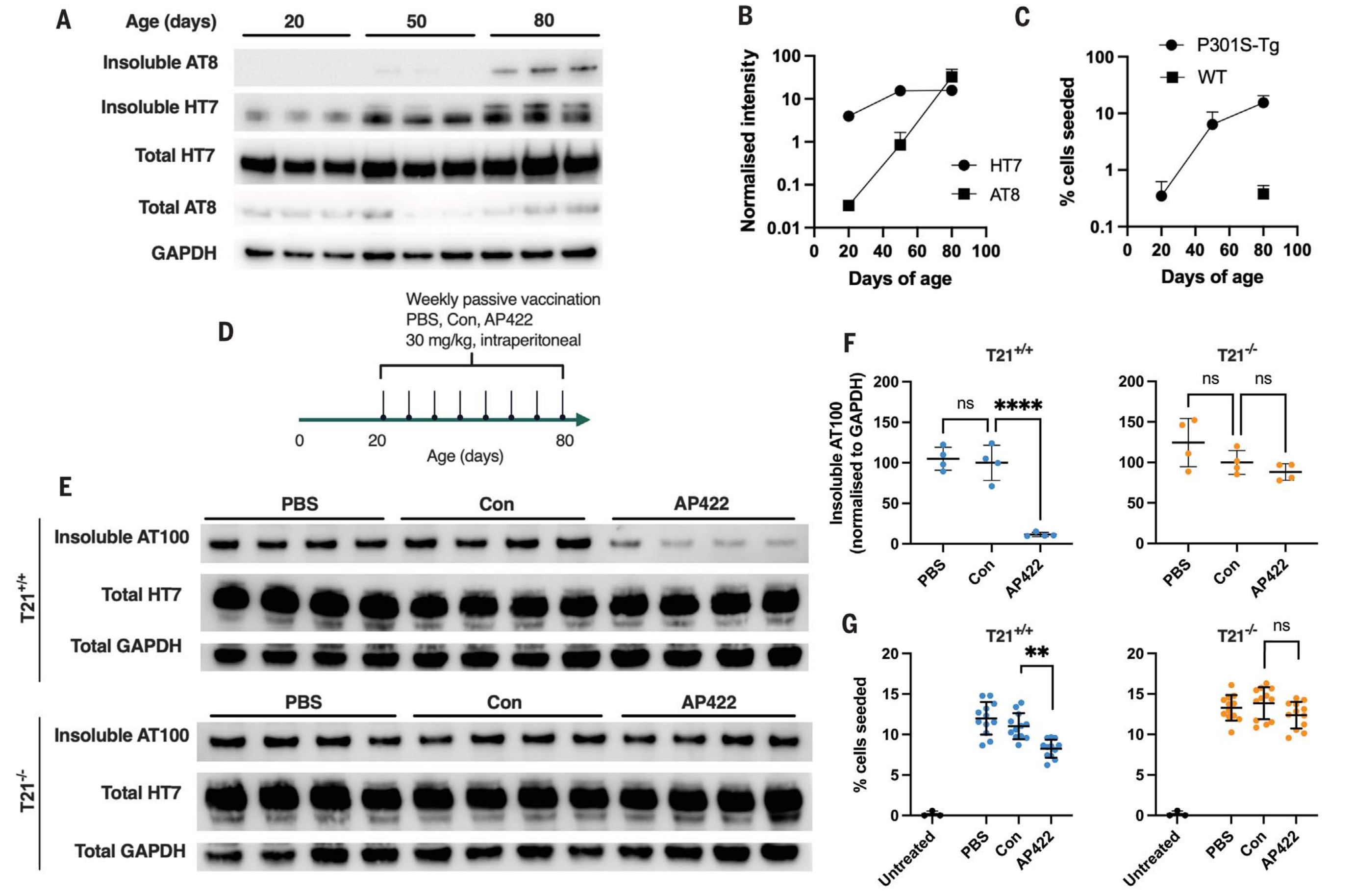


Fig. 3. T21 is required for immunotherapeutic protection against incipient tau pathology. (**A**) Immunoblot analysis of total homogenate and sarkosyl-insoluble fractions prepared from the lumbar spinal column of P301S Tau-Tg mice at postnatal day 20, 50, and 80. Lanes indicate individual animals. GAPDH, glyceraldehyde-3-phosphate dehydrogenase. (**B**) Quantification of tau in sarkosyl-insoluble fractions by using Abs AT8 and HT7, normalized to GAPDH. (**C**) Levels of seeding in HEK293 P301S tau-venus cells treated with the same sarkosyl-insoluble fractions, or with insoluble fractions from wild-type mice. (**D**) Timeline of antibody treatment with mock (PBS), anti-adenovirus 9C12 (Con), or AP422 by weekly intraperitoneal injection between ages 20 and 80 days. (**E**) Immunoblot analysis of total and sarkosyl-insoluble fractions of spines from treated mice. Each lane represents an individual mouse. (**F**) Quantification of AT100 levels normalized to GAPDH from (E). (**G**) Levels of seed-competent tau present in spine sarkosyl-insoluble fraction derived from mice treated with the indicated Ab. Points indidcate average seeding in multiple wells from n = 4 mice. (F) Mean ± SD and one-way ANOVA with Dunnett's multiple comparison test. (G) Mean ± SD and nested one-way ANOVA. $^{**}P < 0.01$; $^{***}P < 0.001$.

mouse spines (Fig. 3C), which indicates that seeding activity arises from transgenic tau. We therefore assessed whether Abs could reduce incipient tau pathology through passive Ab transfer into P301S Tau-Tg T21$^{+/+}$ and P301S Tau-Tg T21$^{-/-}$ mice. Mice were treated with AP422, control Ab 9C12 (Con), or buffer only [phosphate-buffered saline (PBS)] by means of weekly intraperitoneal injection (Fig. 3D). Immunoblot revealed a >85% reduction in insoluble tau levels after AP422 treatment in T21$^{+/+}$ animals (Fig. 3, E and F). However, no Ab protection was observed in T21$^{-/-}$ animals or when control Ab was used. We further examined levels of seed-competent tau species in these preparations and observed a significant T21-dependent reduction after AP422 treatment (Fig. 3G). The protection against sarkosyl-insoluble tau accumulation was of greater magnitude than the reduction in the generation of new seed-competent species. This is of interest because T21 is activated by a stoichiometric threshold of Abs (*40*). Given that tau seeds can be of low stoichiometric value, potentially even monomers (*41*, *42*), our findings suggest that fibrillar tau may represent a better substrate for T21 degradation than small tau assemblies. In summary, Abs can reduce levels of incipient tau aggregation in the mouse brain in a manner that requires T21.

In vivo requirement of T21 during long-term immunotherapy

We further sought to establish the involvement of T21 in a chronic Ab treatment regimen in adult mice. We first examined the persistence of biotinylated Abs in circulation and observed similar half-lives (~7 days) in the P301S Tau-Tg versus P301S Tau-Tg T21$^{-/-}$ mice (Fig. 4A). This suggests that T21 is not involved in determining Ab persistence. We therefore proceeded to treat both genotypes with PBS, 9C12 (Con), or AP422 for 17 weeks by means of weekly administration of Abs to the periphery (Fig. 4B). AP422 conferred potent protection against total sarkosyl-insoluble tau (HT7) and against hyperphosphorylated sarkosyl-insoluble tau species detected by Abs AT8 and anti-pS422, which detects the phospho-epitope targeted by AP422 (Fig. 4, C and D). A reduction in pS422-positive cell bodies was also observed with immunofluorescence microscopy (fig. S7). However, in T21-deficient mice, AP422 was again unable to protect against tau pathology. We quantified the number of seed-competent species in brain homogenate and sarkosyl-insoluble fractions using HEK293 P301S

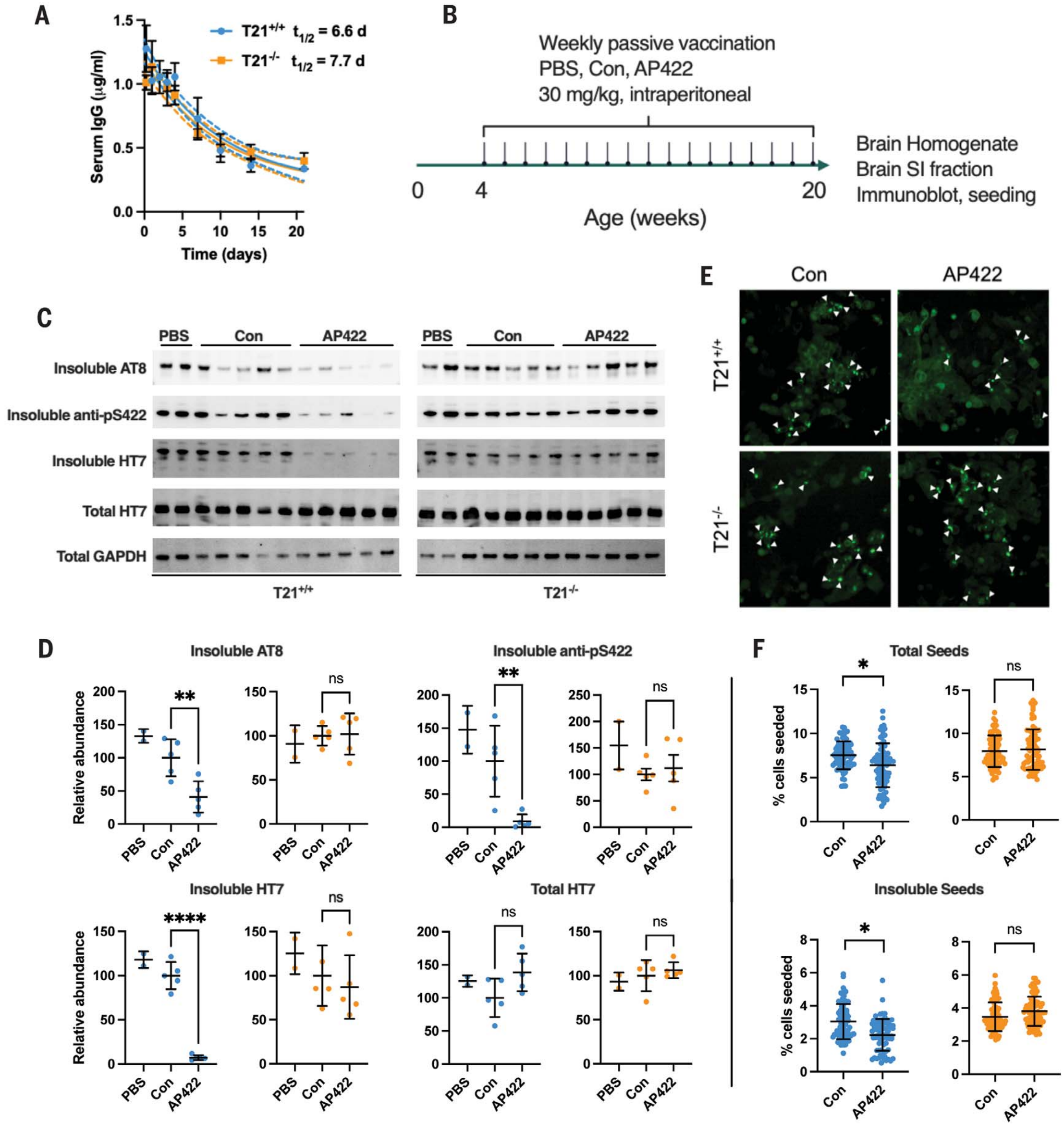


Fig. 4. Long-term immunotherapy potentiates T21-dependent protection against tau pathology. (**A**) Serum concentration of biotinylated Abs at indicated time after injection of 30 mg/kg to $n = 3$ P301S-Tg mice that were either T21$^{+/+}$ or T21$^{-/-}$. (**B**) Cartoon depicting timeline of antibody treatment. Mice were treated with mock intraperitoneal injection (PBS), control Ab 9C12 (Con), or anti-tau (AP422) for 17 weeks. Total homogenate and sarkosyl-insoluble fractions were prepared for immunoblot and quantification of seed-competent species. (**C**) Immunoblot of total and sarkosyl-insoluble fractions from brain hemispheres from P301S-Tg mice that were either T21$^{+/+}$ or T21$^{-/-}$. Samples were probed with either pan-tau monoclonal antibody HT7 or phospho-specific tau Abs AT8 and AP422. Each lane indicates a single mouse. (**D**) Quantification of HT7, AT8, and pS422 levels normalized to GAPDH by using the same samples as (C). (**E**) Images of HEK293 cells that express P301S tau-venus treated with diluted sarkosyl-insoluble fractions from brains treated with the indicated Ab or PBS. (**F**) Quantification of seed-competent tau present in brain homogenates and sarkosyl-insoluble fractions derived from mice treated with the indicated Ab. Points indicate seeding in individual images from $n = 5$ mice. (A) Mean ± SD; one-phase decay curves compared by using extra sum-of-squares F test, not significant. (D) to (F) Mean ± SD. (D) One-way ANOVA with Dunnett's test for multiple comparisons; (F) nested t test for individual mice. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{****}P < 0.0001$.

tau-venus cells. AP422 was able to reduce levels of seeds in both fractions compared with the control 9C12, but only when T21 was expressed (Fig. 4, E and F). Thus, T21 is required for protection against tau pathology during chronic immunotherapy in mice.

Discussion

In this study, we have demonstrated that immunotherapy against tau relies predominantly on the intracellular Ab receptor T21. Mice lacking expression of T21 were refractory to passive immunotherapy at both an early stage of tau pathogenesis and during prolonged Ab treatment. In T21$^{-/-}$ animals, Abs were unable to reduce levels of insoluble tau in the brain and failed to reduce the generation of seed-competent species. In ex vivo slice culture experiments, we observed that seeded aggregation was susceptible to Ab neutralization in a manner that relied largely on T21. We found that other effector mechanisms mediated by cell surface FcγRs, which are expressed widely on microglia in the brain, were less effective than T21 in our models. We found that tau-Ab complexes were internalized to neurons and were recognized by T21 in the cytosol. Our findings are therefore consistent with Abs binding to extracellular tau species before uptake of the tau-Ab complexes to cells. Subsequent binding of T21 to Abs neutralized seeding activity in a manner that required active ubiquitination machinery. Our findings have implications for immunotherapy as a putative treatment in neurodegenerative disease. Human iPSC-derived neurons were found to express regulatable and functional T21. Last, it is possible that proteins other than tau with similar prion-like behavior, such as α-synuclein and TAR DNA-binding protein 43, may be similarly susceptible to T21-mediated neutralization.

REFERENCES AND NOTES

1. M. Goedert, *FEBS Lett.* **592**, 2383–2391 (2018).
2. J. N. Rauch *et al.*, *Sci. Rep.* **8**, 6382 (2018).
3. J. N. Rauch *et al.*, *Nature* **580**, 381–385 (2020).
4. J. M. Cooper *et al.*, *J. Biol. Chem.* **296**, 100715 (2021).

5. B. B. Holmes *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3138–E3147 (2013).
6. F. Clavaguera *et al.*, *Nat. Cell Biol.* **11**, 909–913 (2009).
7. X. Chai *et al.*, *J. Biol. Chem.* **286**, 34457–34467 (2011).
8. A. A. Asuni, A. Boutajangout, D. Quartermain, E. M. Sigurdsson, *J. Neurosci.* **27**, 9115–9129 (2007).
9. K. Yanamandra *et al.*, *Neuron* **80**, 402–414 (2013).
10. M. Bi, A. Ittner, Y. D. Ke, J. Götz, L. M. Ittner, *PLOS ONE* **6**, e26860 (2011).
11. S.-H. Lee *et al.*, *Cell Rep.* **16**, 1690–1700 (2016).
12. W. Luo *et al.*, *Sci. Rep.* **5**, 11161 (2015).
13. K. E. Funk, H. Mirbaha, H. Jiang, D. M. Holtzman, M. I. Diamond, *J. Biol. Chem.* **290**, 21652–21662 (2015).
14. C. R. Andersson *et al.*, *Sci. Rep.* **9**, 4658 (2019).
15. L. D. Evans *et al.*, *Cell Rep.* **22**, 3612–3624 (2018).
16. M. Albert *et al.*, *Brain* **142**, 1736–1750 (2019).
17. J. Gu, E. E. Congdon, E. M. Sigurdsson, *J. Biol. Chem.* **288**, 33081–33095 (2013).
18. E. E. Congdon, J. Gu, H. B. R. Sait, E. M. Sigurdsson, *J. Biol. Chem.* **288**, 35452–35465 (2013).
19. L. Collin *et al.*, *Brain* **137**, 2834–2846 (2014).
20. T. Katsinelos, B. J. Tuck, A. S. Mukadam, W. A. McEwan, *Front. Immunol.* **10**, 1139 (2019).
21. D. L. Mallery *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19985–19990 (2010).
22. W. A. McEwan *et al.*, *J. Virol.* **86**, 8482–8491 (2012).
23. A. J. Fletcher, D. L. Mallery, R. E. Watkinson, C. F. Dickson, L. C. James, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10014–10019 (2015).
24. D. Clift *et al.*, *Cell* **171**, 1692–1706.e18 (2017).
25. W. A. McEwan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 574–579 (2017).
26. A. Kondo *et al.*, *Nature* **523**, 431–436 (2015).
27. M. Goedert, M. G. Spillantini, R. Jakes, D. Rutherford, R. A. Crowther, *Neuron* **3**, 519–526 (1989).
28. B. J. Tuck *et al.*, *Cell Rep.* **39**, 110776 (2022).
29. L. C. James, A. H. Keeble, Z. Khan, D. A. Rhodes, J. Trowsdale, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6200–6205 (2007).
30. L. V. C. Miller *et al.*, *Acta Neuropathol. Commun.* **9**, 41 (2021).
31. R. Yoshimi *et al.*, *J. Immunol.* **182**, 7527–7538 (2009).
32. B. Allen *et al.*, *J. Neurosci.* **22**, 9340–9351 (2002).
33. M. L. Hyer *et al.*, *Nat. Med.* **24**, 186–193 (2018).
34. M. Hasegawa *et al.*, *FEBS Lett.* **384**, 25–30 (1996).
35. M. Lo *et al.*, *J. Biol. Chem.* **292**, 3900–3908 (2017).
36. M. Bottermann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 10440–10445 (2018).
37. W. A. McEwan *et al.*, *Nat. Immunol.* **14**, 327–336 (2013).
38. S. Foss *et al.*, *J. Immunol.* **196**, 3452–3459 (2016).
39. J. A. Macdonald *et al.*, *Acta Neuropathol. Commun.* **7**, 44 (2019).
40. J. Zeng *et al.*, *Nat. Struct. Mol. Biol.* **28**, 278–289 (2021).
41. Z. Hou, D. Chen, B. D. Ryder, L. A. Joachimiak, *Sci. Rep.* **11**, 13602 (2021).
42. H. Mirbaha *et al.*, *eLife* **7**, e36584 (2018).

ACKNOWLEDGMENTS

We thank donors to the Cambridge Brain Bank and their families for their generous gift. We thank M. Hasegawa, Tokyo Metropolitan Institute of Medical Science, Japan, for the provision of antibody AP422 hybridoma and H. C. Morse III, National Institutes for Health, USA, for the provision of *Trim21*-deleted mice. **Funding:** This study was supported by a Sir Henry Dale Fellowship to W.A.M. jointly funded by the Wellcome Trust and the Royal Society (grant 206248/Z/17/Z) and by the Lister Institute for Preventative Medicine. Further support was provided by the UK Dementia Research Institute, which receives its funding from DRI, funded by the UK Medical Research Council, Alzheimer's Society, and Alzheimer's Research UK. A.S.M. was partly supported by Takeda Pharmaceuticals for work unrelated to this project. This work has also received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement 116060 (IMPRiND). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. This work is supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract 17.00038. B.J.T. was supported by the Hughes Hall and Cambridge Trust PhD scholarship. S.S. was supported by a PhD studentship award from Alzheimer's Society, UK (grant AS-PhD-18b-018). L.V.C.M. was supported by UK Medical Research Council DTP PhD scholarship. C.G. was supported by a PhD studentship funded by Alzheimer's Research UK (ARUK-PhD2018-051). The Cambridge Brain Bank is supported by the NIHR Cambridge Biomedical Research Centre (NIHR203312). J.B.R. is supported by the Wellcome Trust (103838 and 220258). S.S.K. was funded by the Lundbeck foundation (R232-2016-2333 and R232-2017-2333). J.T.A., S.F., and S.A.S. were supported by the Civitan Research Fund for Alzheimer's disease, and the Research Council of Norway (grants 287927, 314909, and 335688). L.C.J. and M.V. were supported by a Welcome Trust Investigator Award to L.C.J. (223054/Z/21/Z) and the MRC (UK; U105181010). M.G. is an associate member of the UK Dementia Research Institute and supported by the UK Medical Research Council (MC_U105184291). **Author contributions:** Conceptualization: A.S.M., L.V.C.M., A.E.S., J.T.A., L.C.J., and W.A.M.; Methodology: A.S.M., L.V.C.M., A.E.S., L.C.J., M.G., J.T.A., and W.A.M.; Investigation: A.S.M., L.V.C.M., A.E.S., M.J.V., S.A.S., S.K., S.S., B.J.T., T.K., C.G., and S.F.; Resources: L.S., S.K., S.F., K.M., K.O., M.W., C.K., J.B., E.A., and J.R.; Funding acquisition: J.R., S.S.K., M.G., J.T.A., L.C.J., and W.A.M.; Writing – original draft: W.A.M. and A.S.M.; Writing – reviewing and editing: all authors. **Competing interests:** W.A.M., L.C.J., A.S.M., L.V.C.M., S.K., A.E.S., and B.J.T. are listed as inventors on a patent related to this study. W.A.M. has acted as consultant to Ahren Innovation Capital. J.B.R. has provided consultancy or advisory board support to Astronautx, Astex, Curasen, Biogen, SV Health, WAVE, UCB, Alzheimer Research UK, and LundbeckFonden. **Data and materials availability:** All data are available in the manuscript or the supplementary material.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn1366
Materials and Methods
Figs. S1 to S7
References (*43–49*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 8 November 2021; resubmitted 7 December 2022
Accepted 9 February 2023
10.1126/science.abn1366

RESEARCH ARTICLE SUMMARY

HUMAN GENOMICS

Analysis of genetic dominance in the UK Biobank

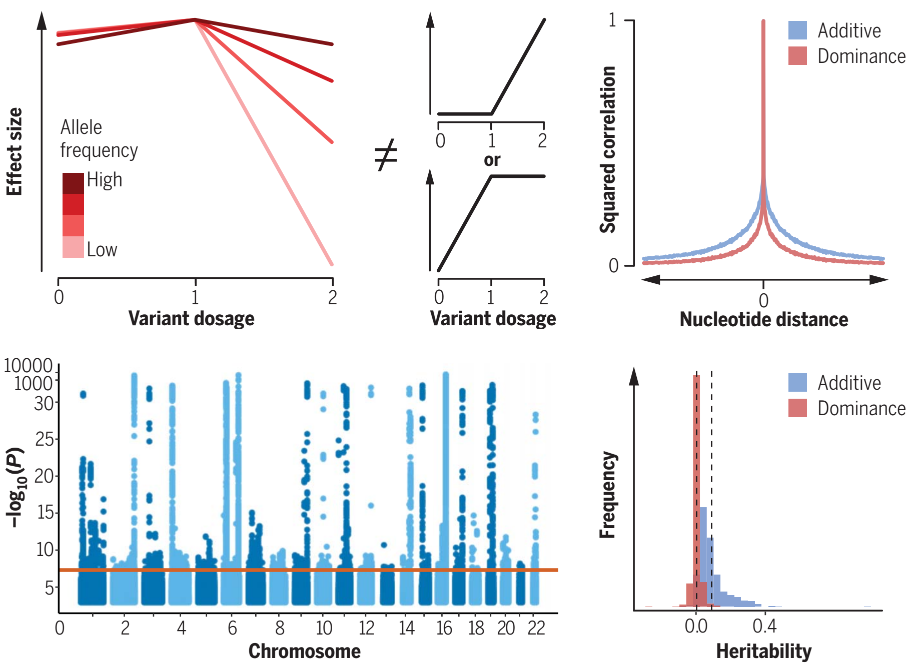
Duncan S. Palmer*, Wei Zhou, Liam Abbott, Emilie M. Wigdor, Nikolas Baya, Claire Churchhouse, Cotton Seed, Tim Poterba, Daniel King, Masahiro Kanai, Alex Bloemendal†, Benjamin M. Neale*†

INTRODUCTION: In statistical genetics, dominance is a deviation from an additive genetic effect on a trait and is well documented in model organisms, particularly in the context of measures of "fitness," and in plant and animal breeding. In humans, however, evidence of nonadditive genetic effects on complex, polygenic traits is sparse. We looked for evidence of nonadditive effects in more than 1000 phenotypes in the UK Biobank population cohort (N = 361,194). To test for "dominance heritability," or the aggregate contribution of nonadditive genetic effects to trait variance genome-wide, we introduce dominance linkage disequilibrium (LD) score regression (d-LDSC). Our method builds upon existing software to include nonadditive effects site by site.

RATIONALE: Identifying nonadditive genetic effects on traits allows us to better understand their underlying biology. Although nonadditive effects are commonly tested in Mendelian disorders, they are rarely tested in human complex traits. Population biobanks allow us to explore questions of genetic architecture at scale and to detect small effect sizes. We sought to test for evidence of nonadditive effects in the UK Biobank. We also investigated whether the less-correlated nature of dominance associations among sites at a locus relative to their additive counterparts can help pinpoint causal variants.

RESULTS: We identified 183 phenotype-locus pairs at genome-wide significance ($P < 4.7 \times 10^{-11}$). We replicated known associations for phenotypes with dominant and recessive patterns of inheritance, for example, hair color at the *MC1R* locus. Qualitatively, we observed stronger nonadditive effects in instances where additive effects are large or the underlying genetic architecture is concentrated in a few loci. The power to detect nonadditive loci was low: We estimate that around a 20- to 30-fold increase in sample size is necessary to capture evidence of dominance effects similar to that observed at additive loci. Applying LDSC and d-LDSC to 1060 traits, we confirmed strong evidence of additive heritability (700 traits, $P < 4.7 \times 10^{-5}$). Despite analyzing a much larger collection of traits with increased power over existing studies, we found little evidence of dominance heritability. We introduced dominance fine-mapping to pinpoint causal variants in the presence of a dominance signal. Gains in fine-mapping resolution due to the rapid decay of dominance LD compared with additive LD are generally outweighed by weaker association signals.

CONCLUSION: We evaluated the contribution of nonadditive genetic effects on trait variation across 1060 traits in the UK Biobank. We identified a modest number of loci and confirmed that heritability explained by dominance is small, in line with previous analyses. Our results support the robustness of the additive model when modeling human complex traits, consistent with the view that most common variants induce small perturbations of continuous latent biological processes aggregated by a mean-field approximation. Furthermore, the additive model typically captures much of the trait variance at a population level, even under classical dominant or recessive patterns. We estimate that for most complex traits, minimum sample sizes of millions are required to detect nonadditive effects at the same strength of association as those reported for additive effects. ■



Summary of the analysis of nonadditive common variant effects site by site and in aggregate across the human genome for more than 1000 traits. The unique recoding of allele counts (up to a linear rescaling) captures any residual genetic association signal at a variant after accounting for the additive effect, known as the dominance deviation (top left). This "dominance encoding" varies with allele frequency, as shown by the color legend. The dominance encoding is generally distinct from biological recessive-dominant architecture, which includes an additive contribution of allelic dosage (top center). The correlation structure between variants under the dominance encoding decays as the square of the additive correlation between variants, as shown by the color legend (top right). The areas under the blue and red curves are average additive and dominance LD scores, respectively. Aggregate Manhattan plots of nonadditive marginal effect sizes assess the dominance deviation across the genome for 1060 traits (bottom left). The y axis is on the $-\log_{10}$ scale up to 30, after which it switches to a $-\log_{10}[-\log_{10}(P)]$ scale to aid presentation. The genome-wide significance threshold ($P < 4.7 \times 10^{-11}$) is displayed in orange. Additive and dominance effect-size estimates coupled with additive and dominance LD scores were used to estimate the additive and dominance contribution to phenotypic variance, known as additive and dominance SNP heritability (bottom right). Additive and dominance SNP heritability estimates across these traits are displayed in blue and red, respectively. The dashed gray lines display the mean estimates across traits.

The list of author affiliations is available in the full article.
*Corresponding author. Email: duncan.stuart.palmer@gmail.com (D.S.P.); bneale@broadinstitute.org (B.M.N.)
†These authors contributed equally to this work.

READ THE FULL ARTICLE AT https://doi.org/10.1126/science.abn8455

HUMAN GENOMICS

Analysis of genetic dominance in the UK Biobank

Duncan S. Palmer[1,2]†*, Wei Zhou[1,2,3], Liam Abbott[1,2,3], Emilie M. Wigdor[4], Nikolas Baya[1,2,3]†, Claire Churchhouse[1,2,3], Cotton Seed[2,3], Tim Poterba[2,3], Daniel King[2,3], Masahiro Kanai[1,2,3], Alex Bloemendal[1,2,3]‡, Benjamin M. Neale[1,2,3]*‡

Classical statistical genetics theory defines dominance as any deviation from a purely additive, or dosage, effect of a genotype on a trait, which is known as the dominance deviation. Dominance is well documented in plant and animal breeding. Outside of rare monogenic traits, however, evidence in humans is limited. We systematically examined common genetic variation across 1060 traits in a large population cohort (UK Biobank, *N* = 361,194 samples analyzed) for evidence of dominance effects. We then developed a computationally efficient method to rapidly assess the aggregate contribution of dominance deviations to heritability. Lastly, observing that dominance associations are inherently less correlated between sites at a genomic locus than their additive counterparts, we explored whether they may be leveraged to identify causal variants more confidently.

The fraction of a trait's variance that can be explained by genetics is defined as its heritability and can be partitioned into additive, dominance, and epistatic contributions. Additive heritability, or narrow-sense heritability, denotes the proportion of a trait's variation that can be captured by additive genetic effects. Dominance heritability is the additional contribution to trait variation that dominance effects (interactions of alleles within a locus) provide over the additive heritability.

There is a wealth of evidence to support that nonadditive effects contribute to phenotypic variance in studies of model organisms, as well as in plant and animal breeding studies (*1*–*8*). Dominance effects on fitness have been extensively studied in model organisms by population geneticists (*9*). In particular, nonadditive effects have been found in traits, including longevity and viability in *Drosophila*, yeast, and *Caenorhabditis elegans*. In these studies, a dominance effect is typically defined as the relative selective advantage of one copy of a mutation to two copies over a wild-type background. Resultant estimates vary widely (*9*–*15*) but suggest that most mutations reduce fitness and are partially recessive. Manna *et al*. (*9*) use the notion of fitness landscapes as a potential explanation; organisms are likely very close to a fitness "peak" because of natural selection, and so a mutation that affects fitness can result in overshooting the peak, leading to nonlinearities.

These observations suggest that traits that are closely related to fitness, such as certain disease traits or traits relating to fertility, have potential underlying dominance effects in humans. Indeed, Yengo *et al*. (*16*) identified a collection of 11 traits in the UK Biobank that display evidence of inbreeding depression, or directional dominance, wherein homozygous alleles are selected against. The authors subsequently estimated the average magnitude of these effects in carefully selected portions of the genome (*17*). They suggested that inbreeding depression likely manifests as a result of many partially recessive deleterious alleles.

Historically, Sewall Wright and R. A. Fisher disputed the origin of dominance effects, with Fisher suggesting modifier variants on large effect loci in contrast to Wright, who suggested that dominance effects are an emergent property of enzymatic biochemical reactions that follow exponential distributions (*18*). Subsequently, gene expression has been suggested as being a nonlinear process under various modeling frameworks (*19*–*21*). Empirically, there is a range of evidence (*22*) for a nonlinear relationship between gene expression and a variety of phenotypes. Examples include the famous threshold effect of the peppered moth in England during the industrial revolution, whereby the heterozygote is sufficient for full expression of a black phenotype (*23*–*25*); heterozygous loss-of-function variants that are tolerated in genes because of sufficient expression of the remaining, unaffected copy (*26*); and interactions between immune genes where certain alleles can exhibit dominant protective effects over risk alleles for autoimmune diseases such as Goodpasture syndrome (*27*). Dominance also has the potential to arise from modifiers that cause allele-specific expression (*28*, *29*).

Identifying nonadditive effects on traits may allow us to better understand their underlying biology. However, nonadditive genetic effects are rarely tested in the analysis of complex traits in humans, despite a landmark paper (*30*) that set the standard for future genome-wide association studies (GWASs) (*31*) and advocated for their reporting in addition to additive effects. Even the simplest model to estimate nonadditive genetic contributions to a trait, Fisher's dominance deviation (*32*), is generally not carried out in GWASs. Nonadditive genetic effects are rarely tested because it is assumed that an additive model captures most of the contribution of a locus to a trait, even in the presence of a dominant architecture, in all but the most extreme cases (*33*).

When nonadditive effects have been examined, outside of rare variants reported to cause recessive Mendelian disease phenotypes (*34*, *35*), their aggregate contribution to trait variation in humans has been estimated to be small. Zhu *et al*. (*36*) estimated that dominance heritability, on average, explains an additional fifth of the variation in a trait over what can be explained by additive effects. Two studies applied to tens of traits in a population biobank of hundreds of thousands of individuals, UK Biobank (*37*), estimated an even lower contribution of dominance heritability to phenotypic variance by using more recent methods (*38*, *39*). Both studies estimated the average dominance heritability as contributing around 1/200 of the additive heritability, with each placing the average dominance heritability at around 0.1%. Estimates from twin studies (potentially capturing the contribution of common and rare genetic variants) are larger: In a study of 10,682 twins, Chen *et al*. (*40*) found that datasets of older (mean ~65 years old) twin pairs resulted in an increased estimate of dominance heritability compared with that in younger twin pairs, with an average dominance heritability of 25% for the 18 traits they examined. Another aggregate study that drew from more than 50 years of twin studies across more than 14.5 million twin pairs estimated that of the ~18,000 traits analyzed, most (69%) were consistent with a simple additive genetic model (*41*).

Population biobanks of hundreds of thousands of individuals offer the opportunity to explore questions of genetic architecture at scale for in more than a thousand traits and detect small effect sizes. In this work, we investigated the role of dominance effects in 1060 phenotypes across 361,194 participants in the UK Biobank through a series of "dominance scans" (*42*). These traits included 71 human disease encodings per International Statistical Classification of Diseases 10th Revision (ICD-10) codes, 81 disease endpoints, and 267 quantitative traits. Additionally, we introduce a rapid method to estimate dominance heritability by extending a commonly used software: linkage disequilibrium (LD) score regression (LDSC) (*43*–*45*). LDSC is a computationally efficient statistical method to estimate the variance in a trait that is attributable to common genetic variation under an additive model. Our method, dominance LD score regression (d-LDSC), extends LDSC to incorporate nonadditive effects site by site,

[1]Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA. [2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [4]Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK.
*Corresponding author. Email: duncan.stuart.palmer@gmail.com (D.S.P.); bneale@broadinstitute.org (B.M.N.)
†Present address: Big Data Institute at the Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford OX3 7LF, UK.
‡These authors contributed equally to this work.

enabling rapid estimation of the dominance contribution to phenotypic variance across thousands of phenotypes.

Testing for dominance effects at each variant

We tested variants for evidence of dominance deviation beyond a purely additive model (Fig. 1A) (*32*). This is a pertinent area of study because the extent of nonadditive contribution to complex traits has yet to be thoroughly examined. Further, such effects may point to relevant biological insights about the nature of variant effects. The phenotypic variance contributions explained by this model are split into two parts in quantitative genetics theory (*32*, *46*): (i) the variance explained by the average effect of allelic substitution (captured by a purely additive model), known as the additive variance, and (ii) the remaining variance (captured by the dominance deviation), known as the dominance variance. We use the terms "nonadditive" and "dominance" interchangeably to refer to effects captured by within-locus interactions that are not explained by the additive model. We refer to the unique recoding of allele counts (up to a linear rescaling) that allows us to test for a dominance deviation directly as the "dominance encoding" (*36*, *38*, *39*, *42*, *47*). This is not the same as recoding allele counts according to canonical biological dominance, in which the presence of at least one copy of the dominant allele explains the variation in the trait at the site ([0, 1, 2] → [0, 1, 1], as illustrated in Fig. 1B). Nor is it the same as an encoding where the heterozygote displays a more pronounced phenotype than either homozygote, known as overdominance (an example of overdominance, [0, 1, 2] → [0, 1, 0], is illustrated in Fig. 1C). In the biological dominance case (Fig. 1B), even if a variant is present in half the population [a minor allele frequency (MAF) of 0.5], the additive model captures most of the variance explained by the locus (88.8%). Furthermore, if a dominance deviation truly exists at a particular genetic variant, then the additional contribution to variance that is explained by the dominance deviation depends on how common that variant is within the population (its MAF). This can be seen by contrasting the top and bottom rows of Fig. 1. Finally, we note that when dominance effects at a locus are considered in GWASs, they are often incorporated by adding a term that encodes the genotypes as [0, 1, 1] or [0, 1, 0] in a joint model with the additive encoding [0, 1, 2] (*48*). The motivation for the dominance encoding chosen here is largely mathematical because it imposes orthogonality on the two components of variance, meaning that we can run the association test in parallel rather than jointly and that the dominance effects estimates will not be contaminated by additive genetic effect spillover (*36*).

Under Hardy-Weinberg equilibrium (HWE), the standardized dominance encoding maps allele counts from [0, 1, 2] to [$-p/(1-p)$, 1, $-(1-p)/p$] for each variant, where p is the MAF of the variant (*36*, *38*, *39*, *42*, *47*). Using this dominance encoding, we assessed within-locus nonadditive effects using dominance GWASs and estimated the relative contribution of additive and nonadditive genetic variation to phenotypic variance in regions that display the strongest signals of association. Figure 1 displays examples of inheritance patterns. In each panel, the variance in the phenotype that is captured by deviations from the best linear fit (shown by the dashed purple line) is exactly the genetic effect that the dominance encoding will estimate. The largest contributions to nonadditive variation in a trait occur when the deviation from additivity is large and the genetic variant is common in the population. Dominance heritability, the variance in phenotype explained by the sum of all of these nonadditive effects, will manifest if biological dominance and overdominance inheritance patterns are widespread at sites that are highly

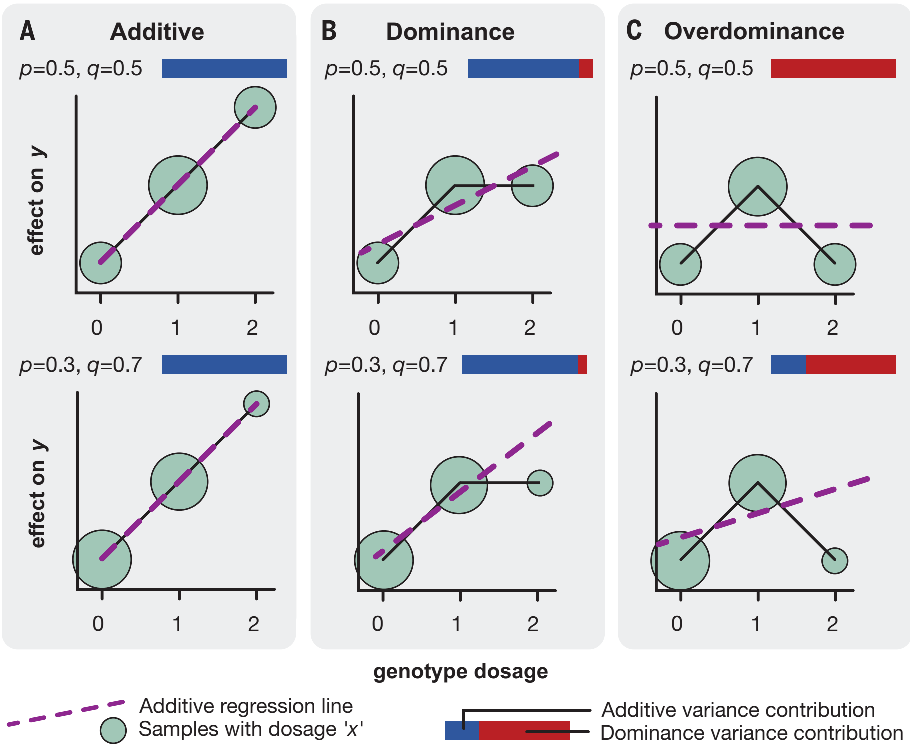


Fig. 1. Examples of inheritance patterns at different MAFs. The effect sizes under a collection of genetic architectures are shown by the black lines. The expected proportions of individuals with each of the three genotypes are shown by a green circle above the alternative allelic dosage (0, 1, or 2); the area of the circle scales with the expected proportion of samples with that genotype. The extra variance captured by deviations from the additive fit (dashed purple line) is the nonadditive or dominance contribution to phenotypic variance. (**A**) Purely additive genetic architecture at the SNP, with no deviation of the truth (black line) from the additive fit (dashed purple line), so no extra variance is explained by nonadditive effects, independent of MAF (top, p = 0.5; bottom, p = 0.3; the variance contribution for both MAFs is entirely blue, representing 100% additive variance contribution at this site). (**B**) Biological dominance architecture at a very common SNP (top, p = 0.5) and at a common SNP (bottom, p = 0.3). Despite being the canonical dominance architecture, additivity explains a large portion of the variance (the dashed purple line is not horizontal, and the variance contribution of additive effects is high; see the length of the blue bar relative to the red bar), but there is an appreciable amount of variation that cannot be explained by a purely additive model. The allele frequency of the SNP matters: The nonadditive variance contribution decreases as MAF decreases from 0.5. Because of the rarity of the homozygous alternate genotype (shown by a smaller green circle beneath the black line at "2" in the bottom graph), the additive model explains a larger portion of the total variance at the SNP. The variance contribution of the recessive contribution is equivalent to the biological dominance encoding of the other allele and amounts to swapping the alternative allele. (**C**) Overdominance at a very common SNP (top, p = 0.5) and at a common SNP (bottom, p = 0.3). When p = 0.5 (top), half the sample is expected to be heterozygous, which completely balances the homozygous individuals so that the additive model explains none of the variance (the dashed purple line is horizontal, and the variance contribution is entirely red) for this genetic architecture at a SNP with p = 0.5. However, overdominance architecture with any other MAF will contain an additive contribution, for example, the bottom graph, where p = 0.3.

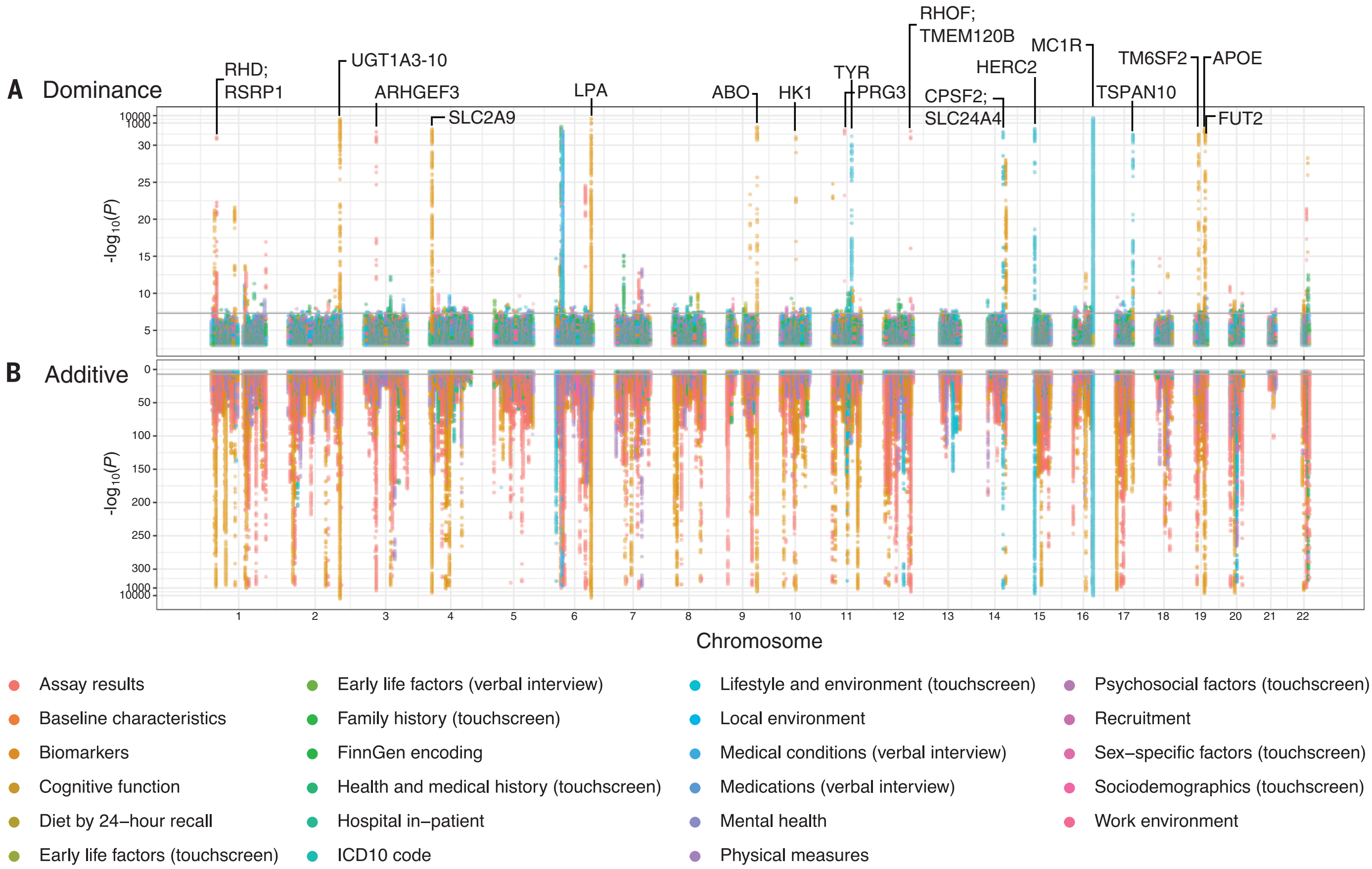


Fig. 2. Aggregate Manhattan plots for additive and nonadditive marginal effect sizes. (**A** and **B**) We determined the most extreme t-statistic for marginal effect sizes across the 1060 traits (*42*) at each site with MAF > 0.05 and plotted the associated $-\log_{10}(P)$ value for the nonadditive (A) and reflected the additive $-\log_{10}(P)$ value (B). Traits are grouped into broad categories as defined in the UK Biobank data showcase and are colored according to the color key. The labels at the top indicate the closest gene to lead associations at top loci ($P < 1.0 \times 10^{-30}$) outside the HLA region. In (A), the y axis is on the $-\log_{10}$ scale up to 30, after which it switches to a $-\log_{10}[-\log_{10}(P)]$ scale. In (B), the y axis is on the $-\log_{10}$ scale up to 300, after which it switches to a $-\log_{10}[-\log_{10}(P)]$ scale. These rescalings of the y axis are to aid presentation. Plots on the $-\log_{10}(P)$ scale are shown in fig. S11.

variable in the population. See fig. S1 for examples of the additive and dominance encoding at varying allele frequency and effect size and how they combine to represent any inheritance pattern at a single-nucleotide polymorphism (SNP).

The dominance encoding may facilitate the identification of causal variants at regions with nonadditive effects. Because DNA is inherited in long tracts from one generation to the next, genetic variants in close proximity tend to be inherited together. The resultant correlated nature of the genome can be represented by the pairwise correlation between variants, known as LD. This correlation structure not only facilitates the detection of genetic associations with a trait, because causal variants are "tagged" by nearby variation, but also makes it difficult to identify true causal variants. LD behaves differently under the dominance encoding than under the additive encoding: In particular, dominance LD between two variant sites is less correlated than the corresponding additive LD [dominance LD between variant pairs is in fact the square of the additive LD between those variants; fig. S2 and (*36*, *42*, *49*)]. We asked whether we could exploit the less-correlated nature of dominance effects to more readily refine marginal dominance signals of association to likely causal variants, a process known as fine-mapping. We explored this idea by developing a modification to existing fine-mapping software (SuSiE) (*50*) to "dominance fine-map" putative dominance associations.

Association studies in the UK Biobank

After quality control and curation of the phenotypic and genotypic data (361,194 samples, 13.7 million variants, 1060 phenotypes), we ran additive and dominance GWASs [(*42*, *51*, *52*); figs. S3 and S4 and table S1]. Binary traits were chosen such that at least eight samples in both categories were expected to be homozygous down to an allele frequency of 5% (*42*). For continuous traits, we analyzed the inverse-rank normal transformation of the raw phenotype to guard against spurious dominance associations [(*42*); figs. S5 to S7 and tables S2 and S3]. After removing variants out of HWE ($P < 10^{-6}$) and restricting to common variants (MAF > 0.05 for binary traits, and MAF > 0.01 for continuous traits), we identified independent signals of association for each phenotype by defining significant regions, known as loci, as 500-kb windows around nominally significant dominance associations ($P < 5.0 \times 10^{-8}$). After merging loci where windows overlapped, we found 183 phenotype-locus pairs that harbored genome-wide significant associations between dominance-encoded genotypes and phenotypes (using a conservative Bonferroni cutoff: $P < \frac{5\times10^{-8}}{1060} = 4.7 \times 10^{-11}$), hereafter referred to as "genome-wide significant loci." Examples include 10 blood cell traits at the *RHD* locus, hair color before graying at the *MC1R* locus, and four bone mineral–density phenotypes upstream of *WNT16*. Among the phenotype-locus pairs, 137 were associations with continuous traits and 46 were association with binary traits. To check for potential artifacts that influenced genotype calling, we examined a collection of variant quality metrics

(*42*) (figs. S8 to S10). Dominance loci were more confidently genotyped and imputed than random allele frequency–matched loci. A summary of our dominance and additive GWAS results are shown in Fig. 2 and fig. S11. We verified that these results were not due to deviations from HWE (*42*) (figs. S12 to S18). We performed additional dominance GWASs in 30 biomarkers without assuming HWE (*53*, *54*) and found highly concordant *P* values of dominance association strength (figs. S17 and S18). Of the reported biomarker dominance loci, 53 of 54 remained significant ($P < 4.7 \times 10^{-11}$), and the remaining phenotype-locus pair was nominally significant ($P = 1.5 \times 10^{-10}$). At lead variants in dominance loci, estimates of the underlying genetic architecture were enriched for monotonic functions of allele count (141/183 = 77.0%).

We replicated known nonadditive effects, including rs1805007 in *MC1R* for hair color (*55*, *56*) ($P < 1 \times 10^{-5000}$ for red hair; fig. S19), an intronic variant of *HERC2* that functions as an enhancer that regulates *OCA2* expression for hair and skin color (*57*, *58*) (rs12913832, $P = 2.87 \times 10^{-56}$ and $P = 5.33 \times 10^{-191}$ for blond hair and skin color, respectively; figs. S20 and S21), and a nonadditive signal-tagging *ILDR1* for hearing difficulty that is stronger than the additive signal at the locus (additive $P = 4.26 \times 10^{-8}$, dominance $P = 5.79 \times 10^{-13}$; fig. S22). *ILDR1* is a known Mendelian hearing-loss gene (*59*–*61*). The stronger dominance signal reflects the high MAF and putative overdominant contribution to the phenotype. We also observed a genome-wide significant nonadditive association with red blood cell distribution width (rs67002563, $P = 2.99 \times 10^{-11}$; fig. S23) in tight LD with *ITPA*. Notably, this locus did not reach genome-wide significance under additivity ($P = 0.000152$; fig. S23). *ITPA* has been implicated in red blood cell disorders through an autosomal recessive mode of inheritance (*62*–*65*). A strong nonadditive association was also observed for the distribution width of platelets at an intronic variant that is associated with increased expression of *ARHGEF3* (rs1354034, $P = 5.80 \times 10^{-90}$; fig. S24) in platelets (*66*). *ARHGEF3* displays a regulatory role in myeloid cell differentiation in zebrafish (*67*). Although this locus harbors a highly significant additive association for platelet count ($P = 2.23 \times 10^{-748}$; fig. S25) and volume ($P = 1.98 \times 10^{-2041}$; fig. S26), the additive signal was completely ablated for platelet distribution width ($P = 0.825$; fig. S24).

The relative contribution of additive and dominance effects to trait variation at top loci

To probe the relative variance explained by the additive and nonadditive contributions, we first examined their relative contributions across the top loci. We extracted the top-five additive and dominance associations across unique cytobands for each phenotype and assessed the relative contribution of additive to nonadditive effects. The median ratio of the two variance contributions was 20.9 (Fig. 3A). When restricting to additive and dominance associations with $P < 1 \times 10^{-6}$, the median ratio of the variance components increased to 28.9 (Fig. 3B). Therefore, we estimate that millions of samples (~7,500,000) are required to be powered to detect nonadditive effects at the same strength of association as those currently reported for additive effects (*68*). This estimate is a best-case scenario: On average, there is much less dominance variance than additive variance at any particular locus. As a result, the expected amount of statistical noise in the dominance effect-size estimate is greater than that in the additive effect-size estimate. Furthermore, given the reduced power in dominance GWASs, effect-size estimates at dominance loci are more susceptible to "winner's curse" than estimates at additive loci.

Fine-mapping dominance loci

Given that not all dominance GWASs were null (Fig. 2) and that the dominance LD decays at the square of additive LD [fig. S2 and (*36*, *42*, *49*)], we investigated whether we could fine-map nonadditive signals of association. To do this, we simply took the dominance LD matrix and dominance GWAS associations as input into existing fine-mapping software (*42*, *69*) instead of their additive counterparts.

We used SuSiE (*50*) to fine-map the dominance signal at nominally significant ($P < 5.0 \times 10^{-8}$) dominance loci. We then fine-mapped any additive effects at these loci and compared the results. A summary is displayed in Fig. 4. While acknowledging the likelihood of winner's curse in these results [because we restricted to nominally significant dominance loci ($P < 5.0 \times 10^{-8}$)], we observed differences in fine-mapping. Of the additively fine-mapped variant-phenotype pairs with more evidence of being causal under the additive model ($n = 690$; yellow shaded area of Fig. 4), 431 reside in genes (*70*). Of these, 23.6% are additive signals without a dominance component [posterior inclusion probability (PIP) < 0.01] that lie in the same gene as a fine-mapped dominance association (PIP > 0.2). Of particular interest were putatively causal variants that were more confidently fine-mapped by their dominance association ($n = 314$; gray shaded area of Fig. 4) because they represent potentially causal association candidates that are not as confidently implicated through additive fine-mapping. Table S4 summarizes the collection of exonic variants that are more confidently dominance fine-mapped. The nonadditive signal in the *ITPA* locus that is associated with red blood cell distribution width was fine-mapped to rs1127354, a nonsynonymous single-nucleotide variant that predicts drug-induced anemia among patients with chronic hepatitis C virus infection (*71*–*75*). The genome-wide significant nonadditive association ($P = 5.79 \times 10^{-13}$; fig. S22) with hearing difficulty was partially fine-mapped to rs2877561, a synonymous change in *ILDR1* that is associated with age-related hearing impairment (*76*), but did not reach genome-wide significance in that study. We note that rs2877561 is a variant associated with changes in expression and splicing for a large number of genes and tissues (*77*).

Genome-wide dominance

After our dominance scans, we also investigated what proportion of phenotypic variance can be explained by dominance effects. Previous papers have suggested, both through theory and empirically, that dominance heritability is likely small in human complex traits

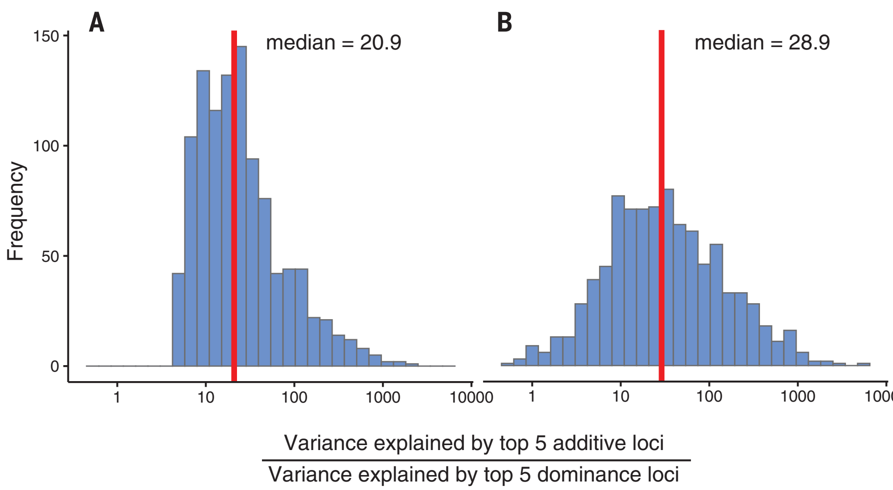


Fig. 3. Relative power to detect dominance associations. (**A** and **B**) The distribution across phenotypes of the ratio of the variance explained by the top-five additive loci to the variance explained by the top-five dominance loci. The *x* axis is on the log scale. In (A), we place no *P* value restriction for inclusion of the least-significant association in the cytoband. In (B), we enforce that each association must have $P < 1 \times 10^{-6}$.

(*33*, *36*). More recently, additional methods to estimate dominance heritability were developed and applied to 50 (*38*) and 70 (*39*) continuous phenotypes in the UK Biobank. These works found zero or marginal evidence that nonadditive effects contribute meaningfully to phenotypic variance genome-wide.

We estimated dominance heritability by extending LDSC. LDSC is a statistical genetics approach that enables computationally efficient estimation of the additive heritability of a trait by relating effect-size estimates of SNPs from GWASs to the extent to which these SNPs tag other SNPs in the genome through LD (their so called LD score). By generalizing the additive model on which the original LDSC software is based to include dominance effects at each variant, we were able to estimate the dominance heritability of traits rapidly, which enabled analysis of thousands of binary and continuous traits within a few minutes (*42*). Indeed, after a dominance GWAS and generation of "dominance LD scores," dominance SNP heritability estimates can be obtained as quickly as additive SNP heritability estimates (*78*). This efficiency allowed us to estimate the nonadditive variance contribution to all 1060 curated phenotypes at low time and economic cost. We performed extensive simulation studies with varying genetic architecture and case-control ascertainment: Dominance heritability estimates were unbiased and well calibrated under all simulation scenarios [(*42*); figs. S27 to S35 and tables S5 and S6).

Dominance heritability of traits in the UK Biobank

Applying additive LDSC and d-LDSC to the 1060 traits, we found strong evidence of significant additive heritability (700 traits with $P < 4.7 \times 10^{-5}$) as expected but little evidence of dominance heritability (Fig. 5 and tables S7 and S8). These findings were robust to the assumed allele frequency dependence on effect size (*42*) (fig. S36). This contrast was present in both continuous and binary traits. For binary phenotypes, we see increased evidence of trait variance that is explained by additive effects at common variants (MAF > 0.05) as case count increases. This trend was not apparent for dominance heritability tagged by common variation. Our results support existing evidence for the modest additional contributions of a nonadditive model to phenotypic variance tagged by common variation over a purely additive model (*36*, *38*, *39*). Across all 1060 curated phenotypes, we found marginal evidence of a small nonzero relative contribution of dominance heritability estimates to additive heritability estimates (York regression; gradient = 0.0023, $P = 0.050$; Fig. 5A). We observed similar results when using a denser set of SNPs down to a MAF of 0.01 to estimate additive and dominance heritability (York regression; gradient = 0.0028, $P = 0.0017$; fig. S37).

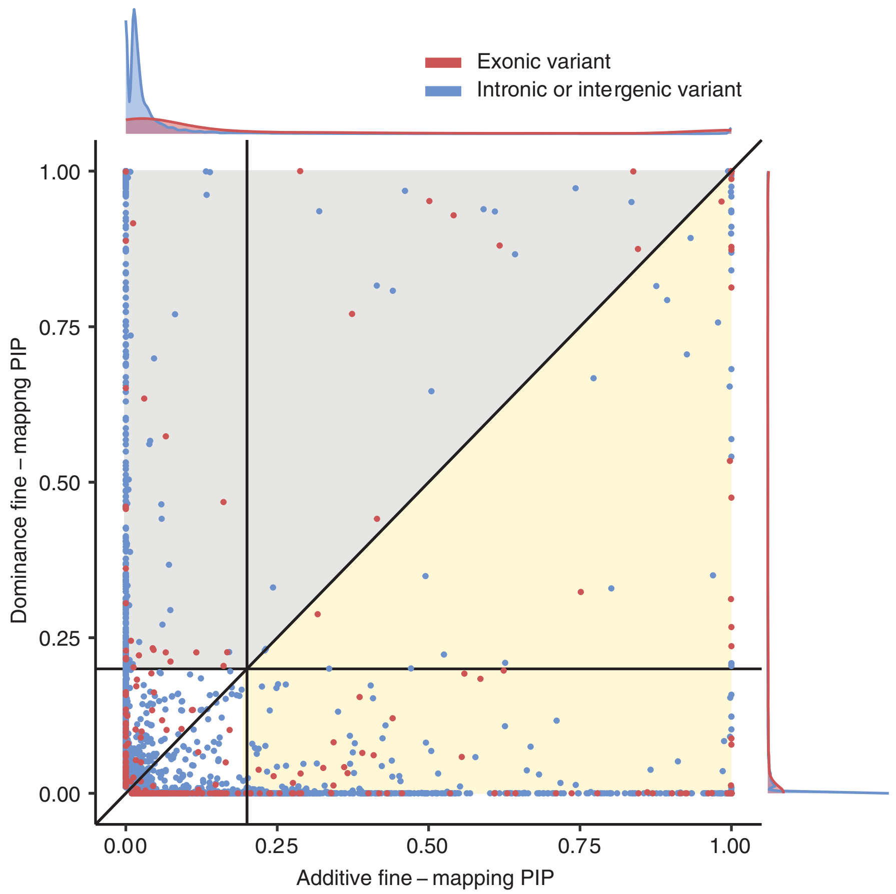


Fig. 4. Fine-mapping dominance loci using SuSiE. We took the collection of genome-wide significant dominance loci ($P < 5.0 \times 10^{-8}$) tagged by SNPs with MAF > 0.05 and fine-mapped using SuSiE (*42*, *50*). This amounted to passing dominance effect sizes and within-sample dominance LD. We then plotted the additive and dominance posterior inclusion probabilities against each other for all dominance loci across all phenotypes. Red points are in the exome, and blue points are intronic or intergenic. The yellow shaded region is the space where additive PIP > dominance PIP and additive PIP > 0.2. The gray shaded region is the space where dominance PIP > additive PIP and dominance PIP > 0.2. Black lines are included to delineate regions. Frequency distributions of additive and dominance PIP are displayed in the margins of the plot.

Although the relative contribution of dominance heritability to additive heritability is low, this does not preclude the possibility that nonadditivity is present or even widespread throughout the genome (Fig. 1) or suggest that nonadditive effects should be disregarded. Nonadditive loci may have large effects on an individual level but may not contribute greatly to the heritability of a trait in the population. Finally, the power to detect deviations from additivity is weakest precisely where we expect the largest nonadditive effects: rarer variation.

Discussion

We performed a large and comprehensive dominance scan and heritability analysis of 1060 phenotypes in the UK Biobank and identified 183 phenotype-locus pairs at genome-wide significance ($P < \frac{5\times10^{-8}}{1060} \approx 4.7 \times 10^{-11}$). These loci consisted of many well-known associations in phenotypes with dominant and recessive patterns of inheritance (for example, hair color). Qualitatively, we observed stronger nonadditive effects in instances where additive effects are large or the underlying genetic architecture is concentrated in a few loci; examples include blood traits, hair color, and biomarkers. For most traits, far more samples are likely necessary to capture evidence of dominance effects. Extrapolating from the top-five loci in each trait, we estimated that millions of samples would be required to obtain marginal dominance effect-size estimates with strengths of association similar to those now observed in additive GWASs.

Despite analyzing a much larger collection of traits with increased power over existing studies (*36*, *38*–*40*), we found limited evidence of a dominance contribution to phenotypic variance. Across the analyzed traits, the mean additive and dominance heritabilities (averaged on the liability scale for binary traits) were 0.088 and 0.00076, respectively. A dominance contribution of around 1/120 of the additive contribution is broadly in line with recent

estimates of 1/200 for the traits analyzed by Pazokitoroudi *et al.* (*38*) and Hivert *et al.* (*39*). We hypothesize that nonsignificant estimates of dominance heritability are due to limited power, owing to the low relative magnitude of dominance variance to additive variance under the most biologically plausible genetic architectures (Fig. 1) (*79*). Our results provide further evidence to support the robustness of the linear model when modeling human complex traits, which reflects that most common variant effects are largely small perturbations of continuous latent biological processes aggregated by a mean-field approximation.

Yengo *et al.* (*17*) introduced a complementary but distinct summary statistic–based approach to ours to estimate inbreeding depression, in which a mean dominance effect size across genetic markers is estimated within an LDSC framework. The authors found enrichment of inbreeding depression within genomic regions with low LD, regions conserved across species, regulatory elements, and regions with an increased contribution to additive heritability. Note that this is a distinct estimator in which the mean nonadditive effect size is estimated, whereas we, as in Pazokitoroudi *et al.* (*38*), constructed an estimator of the average variance contribution of nonadditive effect across the genome assuming a mean effect size of zero.

We introduce dominance fine-mapping to attempt to pinpoint causal variants in the presence of a dominance signal. With the same strength of association as an additive signal, a dominance signal will fine-map more easily. However, gains in fine-mapping accuracy due to the far-more-rapid decay of dominance LD compared with additive LD are generally outweighed by a larger additive signal of association at the locus. A natural extension of this work would be to explore the use of both additive and dominance association signals jointly to increase fine-mapping precision.

There are caveats and limitations to this work. First, throughout this study, we have assumed HWE in the determination of the dominance encoding and subsequent evaluation of dominance effect sizes and heritability. If this assumption is violated, additive effects will be partially captured by the dominance encoding and manifest as nonzero dominance effect sizes. To counter this effect, we imposed a stringent filter to remove variants out of HWE ($P < 1.0 \times 10^{-6}$). In doing so, we may have removed a subset of SNPs under selection with putatively large effects, which we would have been most well-powered to detect. However, we expect that common variants that exhibit this behavior are far more likely to be due to genotyping errors than to true effects of selection. To check the robustness of the HWE assumption, we ran further dominance scans in 30 biomarkers without assuming HWE (*53*, *54*)

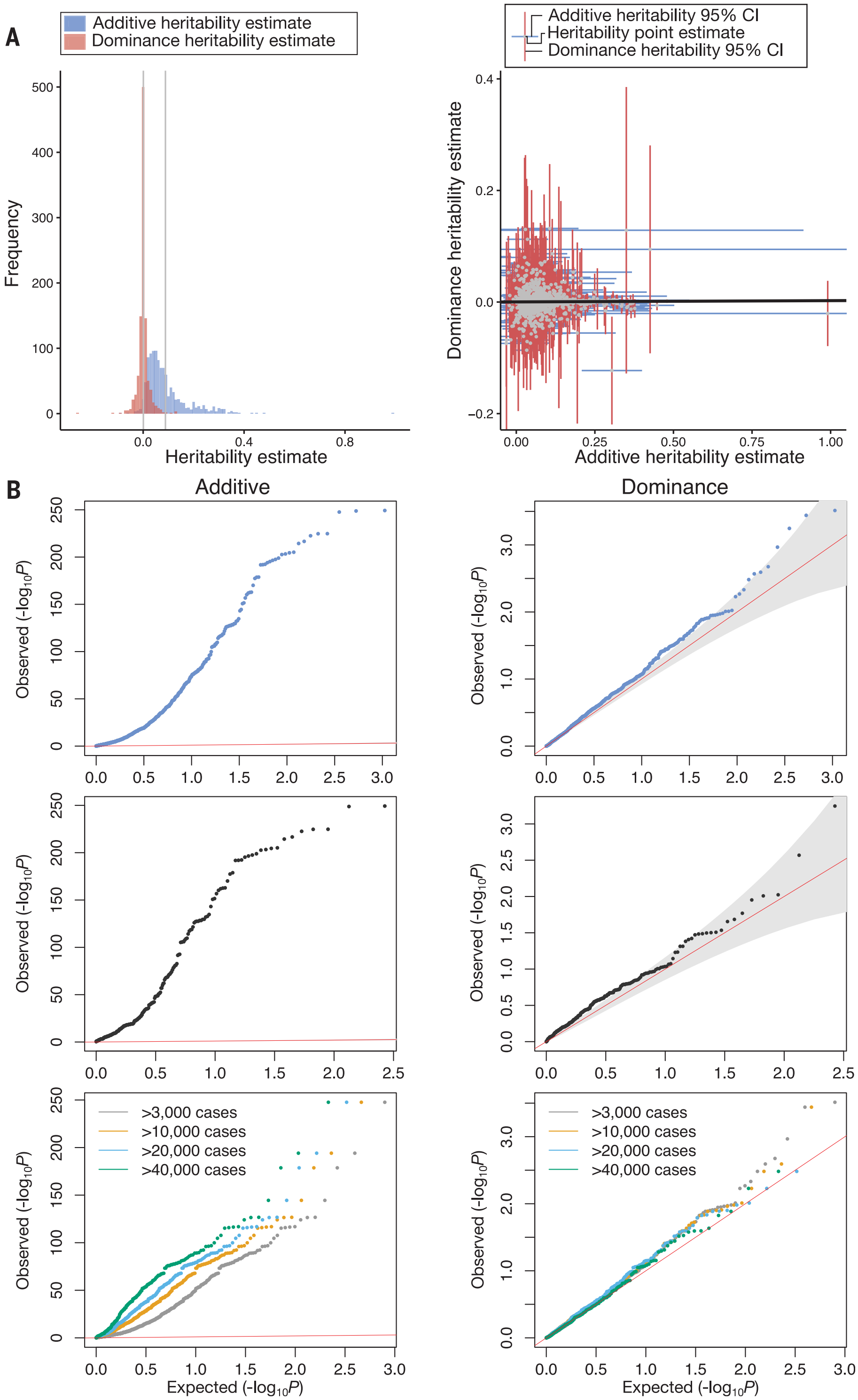


Fig. 5. Summary of heritability analysis. (**A**) Contrasting estimates of additive and dominance heritability for 1060 traits in the UK Biobank. The first column displays histograms of LD score–based estimates of additive and dominance SNP heritability, which are shown in blue and red, respectively. Mean heritability estimates are shown by the gray lines. The second column displays the paired results for each phenotype, colored according to the key. The York regression best-fit line is displayed in black (intercept = 0.00025, gradient = 0.0023). (**B**) Contrasting the statistical evidence for additive and dominance heritability across 1060 traits in the UK Biobank. The first and second columns display quantile-quantile (Q-Q) plots of observed against expected *P* values evaluated using block jackknife standard errors to test if $h^2_A \neq 0$ and $h^2_D \neq 0$, respectively. The top row includes all traits with more than 50,000 data points if continuous or ordinal and more than 3000 cases if binary. In the middle row, we restrict to the continuous and ordinal traits. Finally, in the bottom row, we restrict to the binary traits.

and found highly concordant *P* values of dominance association strength.

Second, in our dominance scans, we assumed the same cutoff for genome-wide significance as for additive GWASs ($P < 5.0 \times 10^{-8}$). However, given the increased number of effectively independent sites in the genome implied by the less-correlated nature of dominance LD, this assumption should be challenged. The benefit that dominance LD provides for fine-mapping is a drawback for GWASs and the detection of phenotypic variance explained by nonadditive effects genome-wide: Dominance LD tagging does not extend as far in the genome as additive LD tagging. Future studies should investigate bottlenecked populations such as those of Finland and Iceland, because many globally rare variants (where we expect the largest dominance effects) will be more common and thus more easily detectable. Moreover, longer haplotypes (i.e., less decay in LD) in these populations may offer enhanced power to detect nonadditive effects at a locus. Third, as we concentrate our analyses of nonadditive effects to common genetic variation, our results do not capture the full spectrum of genetic variation in humans.

Finally, the entirety of our analysis was applied to samples within the UK Biobank of British and Irish ancestry. We made this restriction for two reasons. First, filtering to an ancestry group avoids deviations from HWE that would be induced through sampling from a mixture of ancestry groups. Second, the British and Irish ancestry subset is by far the largest among the homogeneous ancestry groups within the UK Biobank. The generation of increasingly large genetic datasets from non-European ancestry groups will allow us to explore the nonadditive genetic contribution to phenotypes of interest. However, based on our and others' (*36*, *38*, *39*) results, cohorts with orders-of-magnitude-larger sample sizes than are now available are necessary to robustly identify dominance associations in non-European populations, with the exception of consanguineous and bottlenecked populations.

In this work, we estimated dominance heritability tagged by common SNPs genome-wide and did not consider partitions of the genome due to the general paucity of variance explained by nonadditive effects. However, given a nonzero dominance variance, the entirety of the LD score toolkit, including partitioned heritability estimation (*45*) and genetic correlation estimation (*43*), can be applied to dominance effects. In addition to site-by-site dominance effects, LDSC is readily extendable to test gene-by-environment interactions.

We systematically evaluated the contribution of nonadditive genetic effects on trait variation across 1060 traits in the UK Biobank. We found a modest number of individually significant loci and broadly confirmed that heritability explained by dominance is small, in line with previous analyses.

Materials and methods summary

Parametrization of nonadditivity at a locus

The dominance encoding maps the genotypes [0, 1, 2] to $[-p/(1-p), 1, -(1-p)/p]$ for each variant (*36*, *38*, *39*, *42*, *47*). The model we assume, which incorporates an additive effect and dominance deviation, is

$$y = X_A\beta_A + X_D\beta_D + \varepsilon \qquad (1)$$

where y is a phenotype vector across samples; X_A and X_D are matrices of the standardized additive and dominance encoding of the genotypes; β_A and β_D are vectors of causal effect sizes with entries sampled independently from distributions with mean 0 and variance h_A^2/m and h_D^2/m, respectively, where h_A^2 and h_D^2 are the additive and dominance heritabilities, respectively, and m is the number of variants; and ε is a noise term $\sim N\left[(0, 1-(h_A^2+h_D^2)\right]$. For association tests, we regressed y on both $X_{A,j}$ and $X_{D,j}$, separately, for each variant j.

Phenotype preparation

We curated UK Biobank phenotypes for downstream analysis using PHESANT (*42*, *51*, *52*). ICD10 codes were truncated to two digits, and evaluated phenotypes were curated by FinnGen (*80*). We then perform a series of further phenotype-filtering steps (*42*, *81*) to restrict categorical phenotypes to those with ≥3000 cases and controls and to exclude all ordinal variables (table S1).

Sample and variant quality control

Sample quality control

We filtered to unrelated individuals with low autosomal missingness rates (*37*) and European ancestry using the first six principal components (*42*). We removed UK Biobank participation withdrawals and samples without imputed data, resulting in 361,194 individuals for analysis (fig. S4).

Variant quality control

We subsetted to variants with MAF > 0.1% (after sample quality control), HWE $P > 1 \times 10^{-10}$, and info score > 0.8. A subset of rare variants with protein-truncating or missense consequences were also retained (*42*, *82*). For GWASs, 13.7 million variants were retained (fig. S4). Downstream analyses further filtered these variants as described.

Evaluation of additive and dominance marginal effect sizes

We ran GWASs with hail using linear regression (*83*, *84*) with age, age^2, sex, age × sex, age^2 × sex, and the first 20 principal components as covariates.

Heteroskedastic noise simulation study

We ran a Balding-Nichols model with two populations, 1000 samples, and 10,000 independent markers. We then simulated (*85*) 100% heritable phenotypes under the additive model under infinitesimal and spike and slab (1% causal) genetic architecture and added a noise term with a phenotype-dependent variance (*42*). We then ran dominance GWASs on the resultant phenotype with heteroskedastic noise. Under this scenario, no SNP has a nonadditive effect, so any association is artifactual. We then inverse-rank–normalized the phenotypes and reran the GWASs (figs. S5 and S6).

Deviations from HWE

We assumed HWE in the dominance encoding. If this is not the case at a variant, then the additive and dominance encoding of that variant becomes correlated and an additive signal bleeds into the dominance contribution. We tested the impact of deviations from HWE at a variant, *j*, on dominance associations by varying genotype proportions through an inbreeding coefficient: $P(AA) = (1-F)p^2 + pF$, $P(Aa) = 2pq(1-F)$, and $P(aa) = (1-F)q^2 + qF$, where p is the MAF (frequency of A), $q = (1-p)$, and F is the inbreeding coefficient. We then simulated phenotypes: $y = \beta X_j + \varepsilon$, $\varepsilon \sim N(0, 1-\beta^2)$, where X_j is the standardized additively encoded variant *j*. We then applied the dominance encoding assuming HWE and determined the strength of the dominance association in a sample size of 50,000 with a variant whose additive effects explain 5% of the trait variance. We varied the MAF from 0.05 to 0.5 and the inbreeding coefficient from 0 to 0.05. Similarly, we simulated genotyping error and its impact on dominance associations by varying genotype proportions using a genotyping error parameter ϕ: $P(AA) = p^2 + 2pq\phi$, $P(Aa) = 2pq(1-\phi)$, and $P(aa) = q^2$. ϕ denotes the proportion of heterozygous calls that we incorrectly called as homozygous. We reran the same procedure as above, varying ϕ from 0 to 0.05. The results are displayed in figs. S12 and S13. To determine if such a phenomenon is present in our results, we examined the distribution of $|\hat{\beta}_D|$ in the MAF and F bins (figs. S14 to S16).

Similarity of logistic and linear regression with small effect sizes

We assessed whether our choice to run linear regression throughout association testing materially affected our results by running logistic regression at significant ($P < 5.0 \times 10^{-8}$, using linear regression) dominance loci. *P* values of association were highly correlated between linear and logistic regression at dominance loci (mean Pearson correlation of 0.993).

Fine-mapping

We ran SuSiE (*50*) on summary statistics to perform dominance fine-mapping by passing

in-sample dominance LD matrices and dominance summary statistics to the software. Loci were defined by merging 1.5-Mb neighborhoods around significant ($P < 5.0 \times 10^{-8}$) associations. Lead SNPs that were used to define a locus for fine-mapping had to have a MAF > 0.05, but we passed variants within respective loci with less stringent HWE $P > 1 \times 10^{-10}$ and no MAF cutoff. We evaluated LD matrices at each locus using LDstore (*86*). We allowed up to 10 causal variants per locus and a uniform prior for variant causality. Resultant fine-mapped variants with MAF < 0.01 were removed. We excluded the human leukocyte antigen (HLA) region from fine-mapping.

LD score dominance extension

To estimate SNP-based dominance heritability, h_D^2, we constructed a d-LDSC model in which we regress the chi-squared statistics of dominance GWASs on dominance LD scores. The dominance encoding provides a new flavor of LD (*36*, *49*), $r^{\mathrm{DD}} := \frac{1}{n}X'_{\mathrm{D}}X_{\mathrm{D}}$. We can derive a dominance LD score equation relating dominance summary statistics to dominance LD scores, $l_j^{\mathrm{D}} := \sum_k r_{j,k}^{\mathrm{DD}^2}$. Assuming HWE yields a simple, decoupled analog of the additive LDSC (*44*), then

$$E\left[\hat{\beta}_{\mathrm{D},j}^2 | l_j^{\mathrm{D}}\right] \approx \frac{nh_{\mathrm{D}}^2}{m} l_j^{\mathrm{D}} + 1 \quad (2)$$

where n is the number of samples, m is the number of SNPs, h_{D}^2 is the dominance heritability, and $\hat{\beta}_{\mathrm{D},j}^2$ is the marginal nonadditive effect of SNP j on phenotype y. See (*42*) for full details. As in the original LDSC, we may estimate dominance LD scores using an ancestry-matched reference panel. With access to dominance summary statistics and LD scores, we can estimate h_{D}^2 by the gradient of the linear model in Eq. 2. We used a block jackknife (*44*) to obtain standard errors on our estimates. We filtered to HapMap3 SNPs (*87*) and evaluated additive and dominance LD scores within a 1-centimorgan window.

d-LDSC simulation studies

To ensure that d-LDSC is an appropriately conditioned statistical model, we performed an extensive collection of simulation studies. We considered two simulation scenarios: (i) Fully simulated genotypes and phenotypes. We simulated genotype data using msprime (*88*) from 10 million sites in 10 independent chromosomes, sampled 50,000 individuals, and subsetted to variants with MAF > 5%, which was ~250,000 variants. (ii) Real genotypes and simulated phenotypes. We randomly sampled a subset of samples from the quality-controlled UK Biobank data (10,000, 50,000, and 100,000).

We then simulated phenotype data according to Eq. 1, varying h_{A}^2 and h_{D}^2 (0, 0.05, 0.2), genetic architecture (100% variant causal, 10% causal via spike and slab), and trait type [continuous, liability threshold generated binary trait with (6% in population prevalence, 18% in sample prevalence) and without case ascertainment (6% population prevalence)]. Estimates were unbiased and well calibrated (figs. S27 to S35 and tables S5 and S6). See (*42*) for full details.

REFERENCES AND NOTES

1. M. E. Wolak, L. F. Keller, in *Quantitative Genetics in the Wild*, A. Charmantier, D. Garant, E. B. Kruuk, Eds. (Oxford Univ. Press, 2014).
2. J. S. Bloom, I. M. Ehrenreich, W. T. Loo, T.-L. V. Lite, L. Kruglyak, *Nature* **494**, 234–237 (2013).
3. J. P. R. dos Santos, R. C. C. Vasconcellos, L. P. M. Pires, M. Balestre, R. G. Von Pinho, *PLOS ONE* **11**, e0152045 (2016).
4. W. Huang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15553–15559 (2012).
5. M. S. Lopes, J. W. M. Bastiaansen, L. Janss, E. F. Knol, H. Bovenhuis, *G3* **5**, 2629–2637 (2015).
6. M. M. Monir, J. Zhu, *Front. Plant Sci.* **9**, 627 (2018).
7. E. C. Park, H. R. Horvitz, *Genetics* **113**, 821–852 (1986).
8. M. Pettersson, F. Besnier, P. B. Siegel, O. Carlborg, *PLOS Genet.* **7**, e1002180 (2011).
9. F. Manna, G. Martin, T. Lenormand, *Genetics* **189**, 923–937 (2011).
10. A. García-Dorado, A. Caballero, *Genetics* **155**, 1991–2001 (2000).
11. D. Chavarrías, C. López-Fanjul, A. García-Dorado, *Genetics* **158**, 681–693 (2001).
12. D. Houle, K. A. Hughes, S. Assimacopoulos, B. Charlesworth, *Genet. Res.* **70**, 27–34 (1997).
13. K. Szafraniec, D. M. Wloch, P. Sliwa, R. H. Borts, R. Korona, *Genet. Res.* **82**, 19–31 (2003).
14. L. L. Vassilieva, A. M. Hook, M. Lynch, *Evolution* **54**, 1234–1246 (2000).
15. C. D. Huber, A. Durvasula, A. M. Hancock, K. E. Lohmueller, *Nat. Commun.* **9**, 2750 (2018).
16. L. Yengo, N. R. Wray, P. M. Visscher, *Nat. Commun.* **10**, 3719 (2019).
17. L. Yengo *et al*., *Am. J. Hum. Genet.* **108**, 1488–1501 (2021).
18. S. Wright, *Am. Nat.* **68**, 24–53 (1934).
19. R. A. Veitia, *J. Theor. Biol.* **220**, 19–25 (2003).
20. S. Bottani, R. A. Veitia, *Biol. Rev. Camb. Philos. Soc.* **92**, 953–963 (2017).
21. A. H. Porter, N. A. Johnson, A. Y. Tulchinsky, *Genetics* **205**, 101–112 (2017).
22. S. Billiard, V. Castric, V. Llaurens, *Biol. Rev. Camb. Philos. Soc.* **96**, 2925–2942 (2021).
23. W. Bowater, *J. Genet.* **3**, 299–315 (1914).
24. J. B. S. Haldane, E. B. Ford, *Proc. R. Soc. London Ser. B* **145**, 303–306 (1956).
25. A. E. van't Hof *et al*., *Nature* **534**, 102–105 (2016).
26. D. G. MacArthur *et al*., *Science* **335**, 823–828 (2012).
27. J. D. Ooi *et al*., *Nature* **545**, 243–247 (2017).
28. T. Lappalainen *et al*., *Nature* **501**, 506–511 (2013).
29. J. J. Crowley *et al*., *Nat. Genet.* **47**, 353–360 (2015).
30. Wellcome Trust Case Control Consortium, *Nature* **447**, 661–678 (2007).
31. R. J. F. Loos, *Nat. Commun.* **11**, 5900 (2020).
32. R. A. Fisher, *Trans. R. Soc. Edinb.* **52**, 399–433 (1919).
33. W. G. Hill, M. E. Goddard, P. M. Visscher, *PLOS Genet.* **4**, e1000008 (2008).
34. C. E. G. Amorim *et al*., *PLOS Genet.* **13**, e1006915 (2017).
35. J. X. Chong *et al*., *Am. J. Hum. Genet.* **97**, 199–215 (2015).
36. Z. Zhu *et al*., *Am. J. Hum. Genet.* **96**, 377–385 (2015).
37. C. Bycroft *et al*., *Nature* **562**, 203–209 (2018).
38. A. Pazokitoroudi, A. M. Chiu, K. S. Burch, B. Pasaniuc, S. Sankararaman, *Am. J. Hum. Genet.* **108**, 799–808 (2021).
39. V. Hivert *et al*., *Am. J. Hum. Genet.* **108**, 786–798 (2021).
40. X. Chen *et al*., *Am. J. Hum. Genet.* **97**, 708–714 (2015).
41. T. J. C. Polderman *et al*., *Nat. Genet.* **47**, 702–709 (2015).
42. Materials and methods.
43. B. Bulik-Sullivan *et al*., *Nat. Genet.* **47**, 1236–1241 (2015).
44. B. K. Bulik-Sullivan *et al*., *Nat. Genet.* **47**, 291–295 (2015).
45. H. K. Finucane *et al*., *Nat. Genet.* **50**, 621–629 (2018).
46. D. S. Falconer, T. F. C. Mackay, *Introduction to Quantitative Genetics* (Pearson Education, 1996).
47. Z. G. Vitezica, L. Varona, A. Legarra, *Genetics* **195**, 1223–1230 (2013).
48. S. Purcell *et al*., *Am. J. Hum. Genet.* **81**, 559–575 (2007).
49. B. S. Weir, *Annu. Rev. Genomics Hum. Genet.* **9**, 129–142 (2008).
50. G. Wang, A. Sarkar, P. Carbonetto, M. Stephens, *J. R. Stat. Soc. Series B Stat. Methodol.* **82**, 1273–1300 (2020).
51. D. Palmer, K. Karczewski, astheeggeggs/PHESANT: genebass PHESANT fork release. Zenodo (2022); https://doi.org/10.5281/zenodo.6795217.
52. L. A. C. Millard, N. M. Davies, T. R. Gaunt, G. Davey Smith, K. Tilling, *Int. J. Epidemiol.* **47**, 29–35 (2018).
53. J. M. Álvarez-Castro, O. Carlborg, *Genetics* **176**, 1151–1167 (2007).
54. Z. G. Vitezica, A. Legarra, M. A. Toro, L. Varona, *Genetics* **206**, 1297–1307 (2017).
55. J. L. Rees, *Am. J. Hum. Genet.* **75**, 739–751 (2004).
56. M. Visser, M. Kayser, R.-J. Palstra, *Genome Res.* **22**, 446–455 (2012).
57. M. P. Donnelly *et al*., *Hum. Genet.* **131**, 683–696 (2012).
58. R. A. Sturm *et al*., *Am. J. Hum. Genet.* **82**, 424–431 (2008).
59. G. Borck *et al*., *Am. J. Hum. Genet.* **88**, 127–137 (2011).
60. Y. Liu *et al*., *Sci. Rep.* **7**, 7466 (2017).
61. A. Sırmacı *et al*., *Am. J. Hum. Genet.* **86**, 797–804 (2010).
62. N. E. Burgis, *J. Biomed. Sci.* **23**, 73 (2016).
63. M. T. Handley *et al*., *PLOS Genet.* **15**, e1007605 (2019).
64. S. H. Kevelam *et al*., *Ann. Neurol.* **78**, 649–658 (2015).
65. B. S. Vanderheiden, *Biochem. Genet.* **3**, 289–297 (1969).
66. S. Zou *et al*., *PLOS ONE* **12**, e0178095 (2017).
67. C. Gieger *et al*., *Nature* **480**, 201–208 (2011).
68. Neale lab UK-Biobank GWAS results; https://broad-ukb-sumstats-us-east-1.s3.amazonaws.com/round2/additive-tsvs/{file}, where each {file} download link is available at https://bit.ly/additive-GWAS.
69. M. Kanai, mkanai/finemapping-pipeline. Zenodo (2022); https://doi.org/10.5281/zenodo.6908588.
70. K. Wang, M. Li, H. Hakonarson, *Nucleic Acids Res.* **38**, e164 (2010).
71. E. S. El Desoky *et al*., *Clin. Exp. Pharmacol. Physiol.* **44**, 965–968 (2017).
72. M. El Raziky, N. A. Zayed, A. Abdel Baki, S. A. Mansour, R. M. H. Shahin, *J. Med. Virol.* **89**, 1823–1829 (2017).
73. R. Kozuka *et al*., *J. Gastroenterol. Hepatol.* **32**, 1495–1502 (2017).
74. Z. Liu *et al*., *Medicine* **96**, e7554 (2017).
75. Y. Tanaka *et al*., *Ther. Drug Monit.* **41**, 497–502 (2019).
76. T. J. Hoffmann *et al*., *PLOS Genet.* **12**, e1006371 (2016).
77. T. G. Consortium, GTEx Consortium, *Science* **369**, 1318–1330 (2020).
78. D. Palmer, astheeggeggs/d-ldsc: v0.1. Zenodo (2022); https://doi.org/10.5281/zenodo.6908197.
79. W. Huang, T. F. C. Mackay, *PLOS Genet.* **12**, e1006421 (2016).
80. FinnGen, Clinical endpoints; https://www.finngen.fi/en/researchers/clinical-endpoints.
81. UKB Heritability, Heritability of >4,000 traits & disorders in UK Biobank; https://nealelab.github.io/UKBB_ldsc.
82. W. McLaren *et al*., *Genome Biol.* **17**, 122 (2016).
83. Hail, Hail 0.1; https://hail.is/docs/0.1/.
84. Hail Team, Hail. Zenodo (2022); https://doi.org/10.5281/zenodo.7142970.
85. N. Baya, nikbaya/ldscsim: Stable v0.1. Zenodo (2022); https://doi.org/10.5281/zenodo.6826139.
86. C. Benner *et al*., *Am. J. Hum. Genet.* **101**, 539–551 (2017).
87. D. M. Altshuler *et al*., *Nature* **467**, 52–58 (2010).
88. J. Kelleher, A. M. Etheridge, G. McVean, *PLOS Comput. Biol.* **12**, e1004842 (2016).

ACKNOWLEDGMENTS

We thank all members of the Neale lab for comments and suggestions and thank the remaining members of the Hail team for their code suggestions and continued improvements to the Hail codebase. We thank the participants and leadership of the UK Biobank: This work was carried out under UK Biobank application 31063. **Funding:** This study was supported by National Institutes of Health (NIH) grants R01 CA194393 (B.M.N.), R37 MH107649 (B.M.N.), and R01 MH101244 (B.M.N.). **Author contributions:** D.S.P., A.B., and B.M.N. developed the theory; D.S.P. implemented the d-LDSC pipeline; L.A. and D.S.P. implemented the dominance GWAS pipelines; and M.K. and D.S.P. implemented the dominance fine-mapping pipeline. D.S.P. and W.Z. performed all analyses and produced the figures. C.S., T.P., and D.K. implemented computational tools to enable the analyses. A.B. and B.M.N. supervised the project. D.S.P. and N.B. implemented and ran simulation studies. D.S.P. and A.B. wrote the manuscript. B.M.N., E.M.W., M.K., and C.C. provided valuable edits. **Competing interests:** B.M.N. is a member of the scientific advisory board at Deep Genomics and Neumora and a consultant for Camp4

Therapeutics, Takeda Pharmaceutical, and Biogen. D.S.P. was an employee of Genomics plc. All the analyses reported in this paper were performed as part of D.S.P.'s employment at the Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA, and Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. All other authors declare no competing interests. **Data and materials availability:** Dominance summary statistics and heritability estimates are available for download from Amazon web services at https://broad-ukb-sumstats-us-east-1.s3.amazonaws.com/round2/dominance-tsvs/{file}, where each {file} download link can be extracted from the manifest file at https://bit.ly/dominance-GWAS. We also performed sex-specific analyses, curating phenotypes according to the same pipeline (*42*); table S1, fig. S3, and the associated summary statistics files are also available at https://broad-ukb-sumstats-us-east-1.s3.amazonaws.com/round2/dominance-tsvs/. Additive summary statistics results are documented at https://www.nealelab.is/uk-biobank and are available for download from Amazon web services at https://broad-ukb-sumstats-us-east-1.s3.amazonaws.com/round2/additive-tsvs/ {file}, where each {file} download link can be extracted from the manifest file at https://bit.ly/additive-GWAS (*68*). Code implementing the d-LDSC framework is available at https://github.com/astheeggeggs/d-ldsc and Zenodo (*78*). Code to reproduce our analyses is available at https://github.com/astheeggeggs/d-ldsc-paper. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Materials and methods and references cited therein can be found in the main article online.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn8455
Materials and Methods
Figs. S1 to S37
Tables S1 to S8
References (*89–98*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 12 January 2022; accepted 15 February 2023
10.1126/science.abn8455

Check for updates

PALEONTOLOGY

Theropod dinosaur facial reconstruction and the importance of soft tissues in paleobiology

Thomas M. Cullen[1,2,3], Derek W. Larson[4,5], Mark P. Witton[6], Diane Scott[7,8], Tea Maho[7,8], Kirstin S. Brink[9], David C. Evans[5,10], Robert Reisz[7,8]*

Large theropod dinosaurs are often reconstructed with their marginal dentition exposed because of the enormous size of their teeth and their phylogenetic association to crocodylians. We tested this hypothesis using a multiproxy approach. Regressions of skull length and tooth size for a range of theropods and extant varanid lizards confirm that complete coverage of theropod dinosaur teeth with extraoral tissues (gingiva and labial scales) is both plausible and consistent with patterns observed in living ziphodont amniotes. Analyses of dental histology from crocodylians and theropod dinosaurs, including *Tyrannosaurus rex*, further indicate that the most likely condition was complete coverage of the marginal dentition with extraoral tissue when the mouth was closed. This changes our perceptions about the appearance and oral configuration of these iconic predators and has broad implications for our interpretations of other terrestrial animals with large teeth.

The antorbital region of the cranium plays a number of important roles in the biology of terrestrial vertebrates, including respiration, olfaction, and food capture and manipulation. Most known dinosaurs are herbivorous, and some (ornithischians) show evidence for an expanded rictus that formed a superficially cheek-like structure that covered their relatively small dentition externally, with this being particularly relevant for hadrosaurs and ceratopsians (*1–3*). By contrast, many non-avian theropod dinosaurs are renowned for possessing very large teeth, which has led to reconstructions in both scientific and popular literature since the 1980s that show maxillary dentition protruding from their closed mouths rather than covered by extraoral tissues, as in most terrestrial vertebrates (Fig. 1 and fig. S2) (*4*, *5*). Among the arguments in favor of this interpretation are the relatively large sizes of some theropod teeth and evidence from the dinosaur phylogenetic bracket, where crocodylians, the closest extant dentigerous relatives of dinosaurs, lack extensive extraoral tissues (*6–8*). Some recent research on theropod rostral neurovasculature has argued that direct data and evidence are lacking for extraoral tissue reconstructions (*9*), but rigorous reconstructions of these tissues are important for biological inferences for dinosaurs. Theropod dinosaur teeth have relatively thin enamel, and large theropods retained their dentition for prolonged periods of time (*10*, *11*), potentially exposing them to damaging desiccation and wear (*12*). Here, we use multiple lines of evidence, including dental histology, skull and tooth size regressions, and morphological comparisons, to test alternate hypotheses of theropod facial reconstruction.

In extant reptiles, two major anatomical patterns occur with respect to dentition and extraoral tissues. In crocodylians, about one-quarter of the tooth crown height that extends beyond the labial edge of the maxillary bone is covered by a fleshy gingiva, and the enamel-covered crowns are not covered by labial scales ("lips") (Fig. 2). In extant lepidosaurs, which are more distant reptilian relatives to dinosaurs than crocodiles, the base of the teeth is similarly covered in gingiva; however, the enamel-covered crowns of the teeth are covered externally by labial scales when the mouth

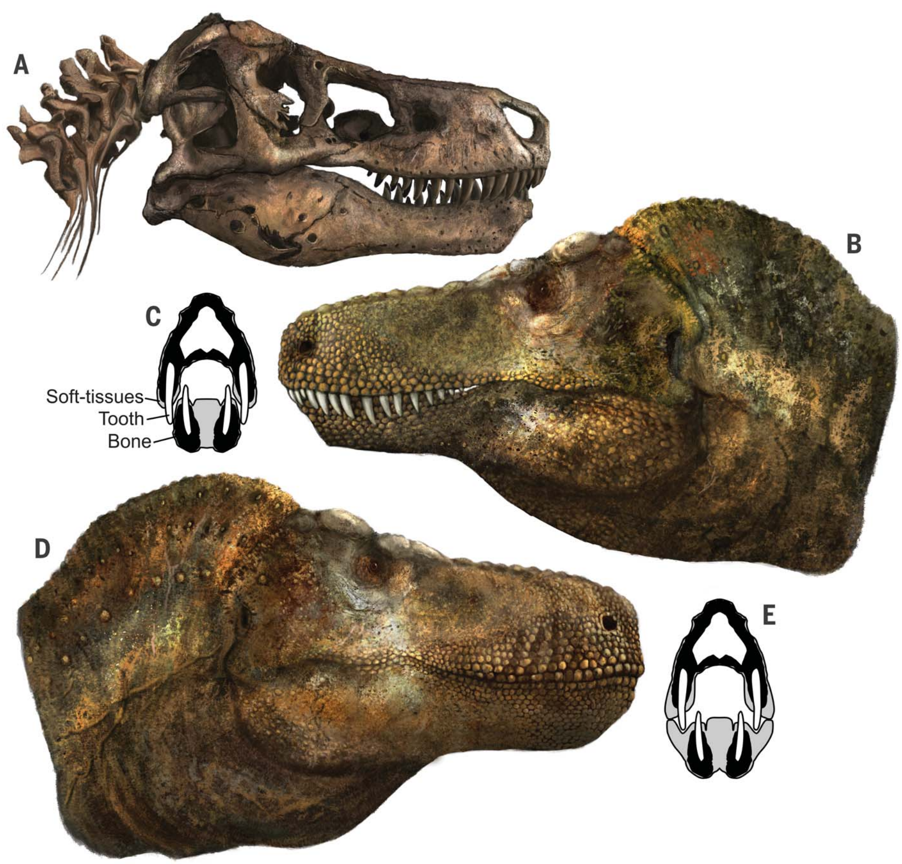


Fig. 1. Comparisons of the reconstructions of *T. rex*. (**A**) Skull, based on Field Museum of Natural History specimen FMNH PR 2081. (**B** to **E**) Two hypothetical flesh reconstructions, one with exposed teeth (B) and an associated cross section of the snout (C) and one with extraoral tissues covering the teeth (D) and an associated cross section of the snout (E).

is closed (Fig. 2). This applies even in large-toothed taxa such as predatory varanid lizards. Notably, in both these lizards and theropod dinosaurs, the teeth are parasagitally aligned with the vertical plane of the skull and do not lean outward as in extant crocodiles (*13*).

Phylogenetic bracketing, in the absence of evidence from birds and fossils, could support the hypothesis that the large teeth of theropod dinosaurs would show the same pattern as that of extant crocodiles, with the upper marginal dentition being exposed when the mouth is closed (*8*). However, such narrow applications of extant phylogenetic bracketing can be problematic when considering dinosaur facial tissues (*3*, *14*, *15*), especially given recent studies into the derived facial integument of crocodylians and its relation to their aquatic lifestyles and sensory adaptations (*16*–*18*). The foramina that are present along the jaw margins of reptiles facilitate the passage of blood vessels and branches of the trigeminal nerve to the extraoral tissues and, in derived crocodylians, house sensory organs that were more widely distributed across the snout (*9*). Extinct terrestrial crocodylomorphs (e.g., the Late Triassic taxon *Hesperosuchus*; Fig. 2) possess a more theropod-like pattern of linearly arranged jaw foramina, as well as ziphodont dentition (Fig. 2) (*7*). Indeed, a broader extant comparison (Fig. 2) demonstrates that the lower-density, linear pattern of foramina on the face and jaws of theropods, such as tyrannosaurids, is as or more similar in structure to that of many extant squamates, such as *Varanus* or *Amblyrhynchus*, than to the pattern observed in extant crocodylians such as *Alligator*. This is concordant with other work suggesting that similarly low densities of linearly arranged facial foramina are a widespread feature in tetrapods that possess extraoral soft tissues (*1*, *9*).

Dentition in reptiles, including dinosaurs, is characterized by the presence of a relatively thin enamel layer that covers the crown of the tooth. Enamel is formed during tooth development through amelogenesis, is not reparable or replaceable, and is invariably thin in most carnivorous reptiles, both fossil and extant (*19*, *20*). In theropod dinosaurs, the thickness of the enamel is similar on the lingual and labial sides of the tooth crown and is somewhat size dependent, with the largest theropod dinosaurs having the thickest enamel (*20*, *21*). Crocodylians generally have overall thicker enamel than dinosaurs, with thicker regions toward the apex of the crown (*21*). In addition, dentine exposure is common in teeth and tusks that are exposed to the environment (*22*). To investigate the dental histology of large theropods in detail, we removed a functional upper maxillary tooth from a large individual of the tyrannosaurid *Daspletosaurus* and examined it for age and enamel ultrastructure in histological thin section under plane-polarized

[1]Department of Geosciences, Auburn University, 2050 Beard Eaves Coliseum, Auburn, AL 36849, USA. [2]Ottawa-Carleton Geoscience Centre, Department of Earth Sciences, Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada. [3]Nagaunee Integrative Research Center, Field Museum of Natural History, 1400 S. Lake Shore Drive, Chicago, IL 60605, USA. [4]Collections Care, Royal BC Museum, 675 Belleville Street, Victoria, BC V8V 9W2, Canada. [5]Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada. [6]School of the Environment, Geography and Geosciences, University of Portsmouth, Burnaby Building, Burnaby Road, PO1 3QL Portsmouth, UK. [7]College of Earth Science, Dinosaur Evolution Research Centre and International Centre of Future Science, Jilin University, Changchun, China. [8]Department of Biology, University of Toronto Mississauga, Mississauga, ON L5L 1C6, Canada. [9]Department of Earth Sciences, University of Manitoba, 125 Dysart Road, Winnipeg, MB R3T 2N2, Canada. [10]Department of Natural History, Royal Ontario Museum, 100 Queen's Park, Toronto, ON M5S 2C6, Canada.
***Corresponding author. Email: robert.reisz@utoronto.ca**

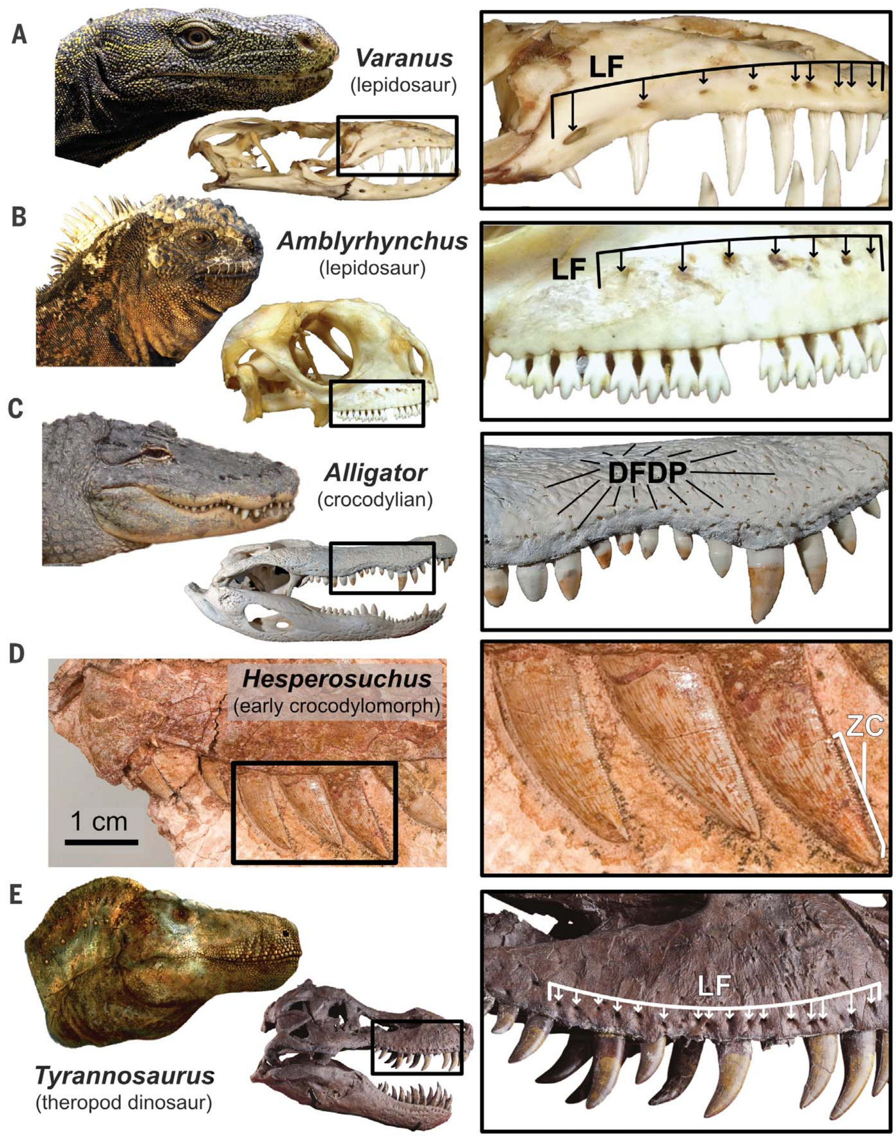


Fig. 2. Comparisons of life appearance and reconstructions, skull shape, and maxillary morphology in lepidosaurs and archosaurs. (**A**) *V. salvadorii*. (**B**) *Amblyrhynchus cristatus*. (**C**) Extant crocodylian *A. mississippiensis*. (**D**) Extinct crocodylomorph *Hesperosuchus agilis*. (**E**) Extinct theropod *T. rex*. Note the linear pattern of foramina (LF) along the extraoral margin in sampled lepidosaurs and theropods in contrast to the broadly distributed pattern of foramina and dome pressure sensor pores (DFDP) in *Alligator*. Also note the ziphodont tooth condition (zc) in the inset image of *Hesperosuchus* (D) compared with the condition present in extant crocodylians. [Image credits: *V. salvadorii*, *A. cristatus*, and *A. mississippiensis* in-life photographs from Wikimedia Commons (public domain); *A. cristatus* skull photograph from E. Graslie (used with permission); *A. mississippiensis* skull photograph from D. Descouens (CC-ASA-4.0); *T. rex* skull photograph from J. Weinstein at FMNH (used with permission); remaining images are from the authors]

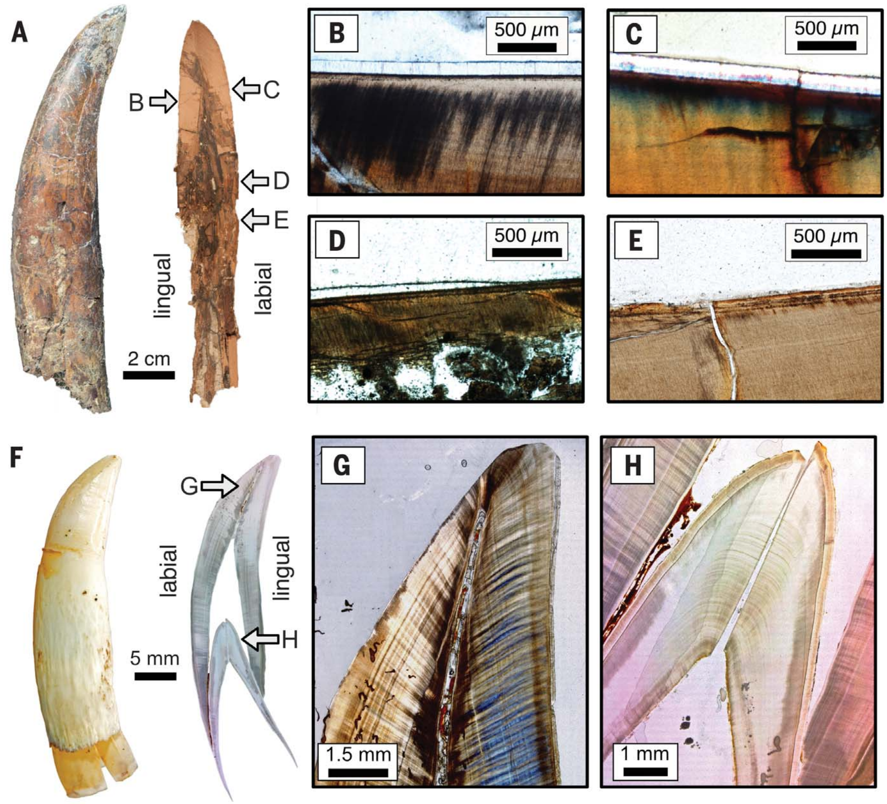


Fig. 3. Histological thin sections of teeth from the large theropod *Daspletosaurus* and *Alligator*. (**A** to **E**) *Daspletosaurus* tooth (Royal Tyrrell Museum of Palaeontology specimen TMP 2003.010.0003) (A) showing relatively unworn enamel of equal thickness on the lingual (B) and labial (C) surfaces of this functional tooth, as well as a reduction of enamel at the base of the crown (D) and cementum present along the root (E). (**F** and **G**) *Alligator* tooth (Royal Ontario Museum specimen ROM R600) (F) showing highly uneven wear patterns between the labial and lingual surfaces (G), with all enamel and some dentine worn away along the labial surface and thick enamel still present on the lingual surface. (**H**) Unerupted *Alligator* tooth without any wear and with the presence of even enamel thickness. See fig. S1 for images of the maxilla of TMP 2003.010.0003 and additional information on the sampled tooth. Images in (B) to (E) and (G) and (H) are thin sections photographed using a petrographic microscope, under plane-polarized [(E) and (H)] and cross-polarized [(B), (D), and (G)] light.

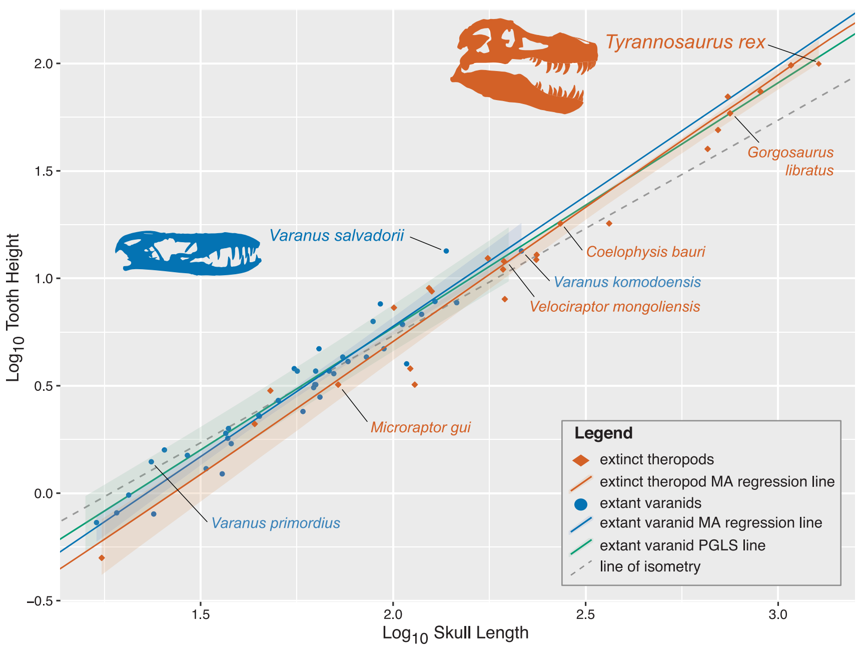


Fig. 4. Plot of $\log_{10}$ skull length to $\log_{10}$ tooth height for a range of extant varanids and extinct theropods. Model II major axis (MA) regressions run on extant varanids (all of which have extraoral tissues covering teeth) (blue points, line, and shaded confidence intervals) and extinct theropods (orange points, line, and shaded confidence intervals). Also plotted are a phylogenetic generalized least squares (PGLS) line for the same extant varanids (green) and the line of isometry (dashed gray). Goodness-of-fit for *Varanus* data is as follows: coefficient of determination (r^2) = 0.9285, and $p < 0.001$. The slope of the *Varanus* MA regression line is 1.215, the slope of the PGLS line is 1.140, and the slope of the theropod MA regression line is 1.218. The *Varanus* and theropod lines are not significantly different ($p = 0.97$).

and cross-polarized light using a petrographic microscope (Fig. 3 and fig. S1; see supplementary text). The thin section confirmed that this tooth was fully developed, with an estimated 512 von Ebner lines being present, consistent with tooth development and replacement rates of well over 1 year that have been estimated in other large tyrannosaurids, including *Tyrannosaurus rex (11)*. The enamel was found to be of similar thickness on both the lingual and labial sides, with no evidence of any substantial wear (Fig. 3, A to E). Despite its advanced age, the tooth still carried well-formed mesial and distal cutting edges (carinae) with delicate serrations [ziphodont (*23*) or incrassate (*24*)]. Wear on tyrannosaurid teeth occurs rarely and primarily on the medial surface of the maxillary dentition because of tooth-on-tooth contact with the opposing dentary teeth (*25*). By contrast, the enamel of the largest teeth of *Alligator mississippiensis*, like the one in tooth position 4 (Fig. 3, F to H), frequently becomes eroded on the exposed labial side, with even a substantial portion of the dentine occasionally worn away.

Enamel has a relatively low water content but is still hydrated and maintained in extant terrestrial vertebrates by glandular secretions in the mouth (*12*, *26*), which arrest detrimental changes in enamel hardness and elasticity (*12*, *27*). Dry enamel has a higher nanohardness and elastic modulus, resulting in stiffer tissue (*12*, *27*), whereas wet enamel is better

at resisting wear and abrasion (*12*). Given the relationship between hydration and wear resistance, and the difficulty of maintaining hydration if a tooth is exposed to air for long periods of time, it is unlikely for functional teeth to remain relatively unworn if exposed, unless the enamel structure and thickness are considerably modified. The comparative lack of wear and abrasion in theropod teeth (Fig. 3, A to E) (*23*), in contrast to the extensive and asymmetric wear [Fig. 3, F to H; see also (*28–30*)] and breakage (*31*) observed through ontogeny in crocodylians, suggests that theropod teeth existed under hydrated conditions consistent with the possession of extraoral tissues.

Although the skulls and teeth of theropod dinosaurs, such as *Daspletosaurus* and *Tyrannosaurus*, are indeed very large compared to those of extant reptiles, major-axis regression analyses demonstrate that the slope of the tooth-skull size relationship in theropods closely matches that observed for extant varanids (Fig. 4), thus refuting interpretations that their teeth were unusually large to the extent that tooth size could preclude extraoral tissue coverage. Even the varanid with the largest relative tooth size (*Varanus salvadorii*) does not have exposed dentition (Figs. 2A and 4), and it possesses greater tooth height–to–skull length ratios (0.096) than the largest sampled theropod, *T. rex* (0.074). These data indicate that theropod teeth were not too large to be covered with extraoral tissues when the mouth was closed and that such a condition would be consistent with what is observed in living amniotes.

The scaling relationships of tooth to skull size between varanids and theropods (Fig. 4) provide further support for the potential inference of soft tissue coverings of the marginal dentition in theropods. Although the relationship between tooth and skull size is weakly positively allometric, the relationship does not greatly affect expected tooth size over the scales represented (dashed versus solid lines in Fig. 4), and data comparing tooth crown height to extraoral tissue height, where available, suggest that a weakly negative allometric-to-isometric relationship exists between these measures (i.e., crown height increases at a slower rate than extraoral tissue height with increasing body size; fig. S4). No change in the presence of complete coverage of teeth with extraoral tissues is noted over a 12-fold increase in size between the smallest and largest *Varanus* skulls in the dataset, despite the inclusion of *Varanus* species with teeth that are proportionally larger (relative to skull size) than those observed in most theropods (e.g., *V. salvadorii*). It would therefore be inconsistent with the data to expect the extraoral tissues to deviate from this pattern over the sixfold size increase between *Varanus komodoensis* and *T. rex*. Given the close fit of multiple lineages of small theropods to the tooth-to-skull size relationship documented in varanids, well-developed extraoral tissues appear likely in smaller members of all major theropod groups, and it is unlikely for tooth height to have exceeded facial soft tissue growth, even in larger theropods.

These comparisons show that extraoral tissues of nonavian theropods (Fig. 1 and fig. S2; see supplementary text) were more like those of extant lepidosaurs and other tetrapods than those of birds or crocodylians and that the faces of extant archosaurs do not accurately reflect the ancestral condition of the archosaurian clade. The results of this study strongly support "lipped" facial reconstructions in theropods with wide-reaching implications for their portrayal in science and popular culture. More importantly, the presence of extensive extraoral tissues has implications for tooth strength, feeding ecology, and biomechanics and therefore may have played an important role in how carnivorous theropod dinosaur teeth resisted forces associated with feeding close to the bone and even may have permitted carcass dismemberment while reducing spalling in large tyrannosaurids. Finally, we posit a lepidosaur-like plesiomorphic condition for extraoral tissues in Dinosauria and expect that our results not only will provide a deeper understanding of the evolution of buccal soft tissues generally and advanced oral processing in ornithischians in particular but also, more broadly, will open new directions of research into the relationships between oral soft tissues and feeding behavior in terrestrial vertebrates with large teeth.

REFERENCES AND NOTES

1. A. C. Morhardt, *Dinosaur Smiles: Do the Texture and Morphology of the Premaxilla, Maxilla, and Dentary Bones of Sauropsids Provide Osteological Correlates for Inferring Extra-Oral Structures Reliably in Dinosaurs?* (Western Illinois Univ., 2009).
2. A. Nabavizadeh, *Anat. Rec.* **299**, 271–294 (2016).
3. F. Knoll, *Neues Jahrb. Geol. Paläontol. Abh.* **248**, 355–364 (2008).
4. G. S. Paul, *Predatory Dinosaurs of the World* (Simon and Schuster, 1988).
5. L. M. Witmer, *Science* **293**, 850–853 (2001).
6. C. A. Brochu, *J. Vertebr. Paleontol.* **22**, 1–138 (2003).
7. C. A. Brochu, *Annu. Rev. Earth Planet. Sci.* **31**, 357–397 (2003).
8. T. D. Carr, D. J. Varricchio, J. C. Sedlmayr, E. M. Roberts, J. R. Moore, *Sci. Rep.* **7**, 44942 (2017).
9. F. Bouabdellah, E. Lessner, J. Benoit, *Palaeontol. Electronica* **25**, 1–20 (2022).
10. G. M. Erickson, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14623–14627 (1996).
11. M. D. D'Emic *et al.*, *PLOS ONE* **14**, e0224734 (2019).
12. J. Zheng *et al.*, *Wear* **301**, 316–323 (2013).
13. B. G. Fry *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8969–8974 (2009).
14. T. M. Keillor, in *Tyrannosaurid Paleobiology*, J. M. Parrish, R. E. Molnar, P. J. Currie, E. B. Koppelhus, Eds. (Indiana Univ. Press, 2013), pp. 157–175.
15. M. P. Witton, *The Palaeoartist's Handbook* (Crowood Press, 2018).
16. M. C. Milinkovitch *et al.*, *Science* **339**, 78–81 (2013).
17. D. Soares, *Nature* **417**, 241–242 (2002).
18. E. J. Lessner, K. N. Dollman, J. M. Clark, X. Xu, C. M. Holliday, *J. Anat.* joa.13826 (2023).
19. P. M. Sander, *Münch Geow. Abhandl. A. Geologie. Palae.* **38**, 87–97 (1999).
20. S. H. Hwang, *Biol. Rev. Camb. Philos. Soc.* **86**, 183–216 (2011).
21. K. Sellers, A. Schmiegelow, C. Holliday, *J. Zool.* **309**, 172–181 (2019).
22. A. Nasoori, *Arch. Oral Biol.* **117**, 104835–104835 (2020).
23. K. S. Brink *et al.*, *Sci. Rep.* **5**, 12338 (2015).
24. J. Lü *et al.*, *Nat. Commun.* **5**, 3788 (2014).
25. B. W. Schubert, P. S. Ungar, *Acta Palaeontol. Pol.* **50**, 93–99 (2005).
26. S. W. Keenan, R. M. Elsey, *Integr. Comp. Biol.* **55**, 972–985 (2015).
27. X. Wang, N. Zhang, Y. Zhong, F. Yan, B. Jiang, *Mater. Sci. Eng. C Mater. Biol. Appl.* **100**, 354–362 (2019).
28. J. Enax *et al.*, *J. Struct. Biol.* **184**, 155–163 (2013).
29. A. R. LeBlanc, K. S. Brink, T. M. Cullen, R. R. Reisz, *J. Vertebr. Paleontol.* **37**, e1354006 (2017).
30. A. R. LeBlanc, R. R. Reisz, D. C. Evans, A. M. Bailleul, *BMC Evol. Biol.* **16**, 152 (2016).
31. G. M. Erickson, *Copeia* **1996**, 739–743 (1996).

ACKNOWLEDGMENTS

We thank K. Chiba and Y. Haridy for assistance with histological thin sections of dinosaur and crocodylian teeth. Access to fossil and extant materials was helpfully provided by K. Seymour [Royal Ontario Museum (ROM)], R. MacCulloch (ROM), A. Lathrop (ROM), N. Richards (ROM), B. Simpson [Field Museum of Natural History (FMNH)], A. Resetar (FMNH), K. Kelly (FMNH), B. Strilisky [Royal Tyrrell Museum of Palaeontology (TMP)], A. Henrici [Carnegie Museum (CM)], D. Kizirian [American Museum of Natural History (AMNH)], R. Pascocello (AMNH), R. Sadlier [Australian Museum (AM)], C. Beatson (AM), J. Rosado [Museum of Comparative Zoology (MCZ)], G. Schneider [University of Michigan Museum of Zoology (UMMZ)], A. Wynn (Smithsonian Institution National Museum of Natural History), G. Watkins-Colwell [Yale Peabody Museum (YPM)], X. Xu [Institute of Vertebrate Paleontology and Paleoanthropology of the Chinese Academy of Sciences (IVPP)], Z. Zhou (IVPP), and C. Sullivan [IVPP (now at University of Alberta)]. We thank J. Weinstein (FMNH) for permission to use photographs of FMNH PR 2081 in this study. **Funding:** Support for this work was provided by Jilin University, the University of Toronto Mississauga, and Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant 2020-04959 (R.R.); a NSERC Canada Graduate Scholarship, the Kenneth C. Griffin Fund, and NSERC Postdoctoral Fellowship PDF-545802-2020 (T.M.C.); NSERC Discovery Grant 2021-00364 (K.S.B.); and Dinosaur Research Institute 2011 and 2015 Student Project Grants (D.W.L.). **Author contributions:** Conceptualization: R.R., D.S., D.C.E.; Methodology: T.M.C., D.W.L., T.M., D.C.E.; Investigation: All authors; Visualization: T.M.C., M.P.W., D.W.L., D.S., T.M.; Writing – original draft: R.R., T.M.C., K.S.B., M.P.W., D.W.L.; Writing – review and editing: T.M.C., K.S.B., D.C.E., M.P.W., R.R. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo7877
Supplementary Text
Materials and Methods
Figs. S1 to S4
Tables S1 to S3
References (*32–68*)
MDAR Reproducibility Checklist
Data S1 and S2

Submitted 6 March 2022; accepted 3 March 2023
10.1126/science.abo7877

Check for updates

NEUROSCIENCE

Orphan receptor GPR158 serves as a metabotropic glycine receptor: mGlyR

Thibaut Laboute[1], Stefano Zucca[1], Matthew Holcomb[2], Dipak N. Patil[1]†, Chris Garza[2], Brittany A. Wheatley[3], Raktim N. Roy[3], Stefano Forli[2], Kirill A. Martemyanov[1]*

Glycine is a major neurotransmitter involved in several fundamental neuronal processes. The identity of the metabotropic receptor mediating slow neuromodulatory effects of glycine is unknown. We identified an orphan G protein–coupled receptor, GPR158, as a metabotropic glycine receptor (mGlyR). Glycine and a related modulator, taurine, directly bind to a Cache domain of GPR158, and this event inhibits the activity of the intracellular signaling complex regulator of G protein signaling 7–G protein β5 (RGS7-Gβ5), which is associated with the receptor. Glycine signals through mGlyR to inhibit production of the second messenger adenosine 3′,5′-monophosphate. We further show that glycine, but not taurine, acts through mGlyR to regulate neuronal excitability in cortical neurons. These results identify a major neuromodulatory system involved in mediating metabotropic effects of glycine, with implications for understanding cognition and affective states.

Glycine is the simplest amino acid ubiquitously present in all mammalian tissues. Glycine serves as an inhibitory neurotransmitter, but it can be excitatory in developing neurons (*1*, *2*). Glycinergic neurons are distributed across the brain; however, glycine can also be released by glial cells (*3*). Known receptors for glycine belong to the family of pentameric ligand-gated ion channels (*4*). Glycine also serves as a coagonist of *N*-methyl-D-aspartate (NMDA) receptors (*5*). Metabotropic neuromodulatory effects of glycine have been observed (*6*, *7*), but no receptors mediating these actions have been found. Glycine has distinct effects on neural circuits (*3*), and glycinergic transmission has been implicated in pathological conditions, including depression (*8*–*10*).

Metabotropic neuromodulation in the nervous system is mediated mainly by heterotrimeric GTP-binding protein (G protein)–coupled receptors (GPCRs). GPCRs play essential roles in neuronal physiology and pathology and present targets for drug development (*11*). Canonically, GPCRs transduce their signals by activating heterotrimeric G proteins (*12*, *13*). However, G protein–independent modes of signal transduction triggered by the recruitment of β-arrestins and other scaffolds to activated GPCRs have also been described (*14*–*16*). G protein signaling is controlled by regulator of G protein signaling (RGS) proteins, which facilitate their deactivation (*17*). RGS proteins also interact with several GPCRs (*18*–*22*).

GPCRs mediate the effects of all major neurotransmitters except glycine and taurine. However, many GPCRs still have no identified endogenous ligands. Orphan GPCRs may have potential for obtaining insights into physiology and for drug development (*23*, *24*).

GPR158 is one of the most abundant orphan GPCRs in the brain that transduces signals by coupling to RGS proteins (*25*, *26*). In neurons, it regulates signaling to the second messenger adenosine 3′,5′-monophosphate (cAMP) and controls key ion channels, kinases, and neurotrophic factors involved in neuronal excitability and synaptic transmission (*25*, *27*). Accordingly, GPR158 has been heavily implicated in cognition and affective states (*25*, *28*, *29*). Genetic suppression of GPR158 in mice results in a prominent antidepressant phenotype and stress resiliency, making GPR158 an attractive target for development of new antidepressants (*25*).

The endogenous ligand for GPR158 remains unknown. Recent structures of GPR158 revealed the presence of an extracellular Cache domain, a putative ligand-binding module (*30*, *31*).

Results

Glycine signals through GPR158 to regulate cAMP

The structure of GPR158 revealed the presence of a Cache domain, which serves as a ubiquitous ligand-binding module in bacterial chemoreceptors (*30*). We found that

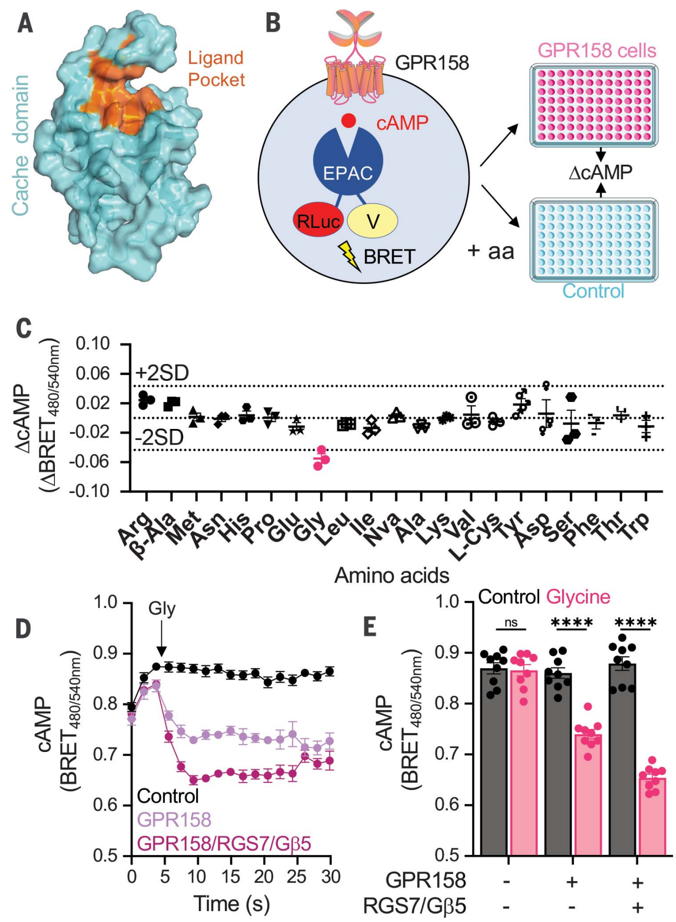


Fig. 1. Identification of glycine as GPR158 ligand. (**A**) Three-dimensional model of the GPR158 Cache domain (cyan) with putative ligand-binding pocket (orange). (**B**) Schematic of the screening assay design. (**C**) Quantification of cAMP changes mediated by GPR158. BRET signal in control cells is subtracted from the signal from cells expressing GPR158, and the difference is plotted. Dotted lines denote 2× SD confidence interval. Data represent mean ± SEM determined from three independent experiments performed in triplicate. (**D**) Time course of cAMP change induced with 10 μM forskolin in U87 MG glioblastoma cells in response to glycine (100 μM) application (arrow). (**E**) Quantification of maximal amplitudes of responses to glycine in (D). Data represent mean ± SEM determined from three independent experiments performed in triplicate. $^{**}p < 0.01$, one-way analysis of variance (ANOVA).

[1]Department of Neuroscience, UF Scripps Biomedical Research, Jupiter, FL 33458, USA. [2]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037 USA. [3]Department of Integrative Structural and Computational Biology, UF Scripps Biomedical Research, Jupiter, FL 33458, USA.
***Corresponding author. Email: kmartemyanov@ufl.edu**
†Present address: Lilly Biotechnology Center, Eli Lilly and Company, 10290 Campus Point Dr., San Diego, CA 92121, USA.

GPR158 Cache domain had a small pocket with organization similar to that of the amino acid binding pocket in other Cache domains (Fig. 1A). We hypothesized that GPR158 may have an amino acid ligand. We screened a library of amino acids for their ability to alter GPR158-mediated signaling. Because GPR158 has been linked to regulation of cAMP in the brain (*25*, *27*), we used a bioluminescence resonance energy transfer (BRET)–based cAMP biosensor (*32*) (Fig. 1B). Out of all amino acids tested, only glycine showed significant decrease in cAMP when applied to human embryonic kidney (HEK) 293 cells expressing GPR158 relative to nontransfected cells (Fig. 1C).

To study this effect in more detail, we analyzed the individual responses to glycine in a kinetic mode. We found that glycine application to U87 glioblastoma cells expressing GPR158 resulted in cAMP decrease. No glycine-induced changes in cAMP were observed in cells lacking GPR158 (Fig. 1, D and E). This inhibitory effect of GPR158 was further potentiated by coexpressing RGS7-G protein β5 (RGS7-Gβ5), suggesting that GPR158 signals by means of this protein complex to affect cAMP levels (Fig. 1, D and E).

We further tested the effect of taurine, a compound closely related to glycine, which binds to several common receptors (*33*), including ionotropic glycine receptors (*34*). Taurine caused a significant decrease in cAMP levels only in HEK293 cells expressing GPR158 (fig. S1, A and B). Again, this effect was potentiated by coexpressing RGS7-Gβ5, suggesting that these proteins act in complex with GPR158 in mediating the effects of taurine. However, when compared directly, the effect of taurine on GPR158-mediated suppression of cAMP was weaker than the effect of glycine (fig. S1C).

Glycine inhibits modulation of RGS7-Gβ5 by GPR158

To understand how glycine action on GPR158 regulates intracellular cAMP, we focused on GPR158 interaction with RGS7-Gβ5, an established guanosine triphosphatase (GTPase)–activating protein (GAP) for the $G\alpha_{i/o}$ proteins (*35*) known to regulate cAMP production (*26*). We used a cell-based assay to monitor GAP activity by following kinetics of G protein deactivation (*36*) (Fig. 2A). In this assay, activation of G proteins by GPCR stimulation generates the BRET signal upon interaction of liberated Venus-Gβγ subunits with the masGRK3CT-Nluc reporter. This signal is quenched when Gα deactivation is triggered by GPCR antagonism and recombines with Venus-Gβγ to form inactive heterotrimer. As previously reported (*22*), we found that introduction of RGS7-Gβ5 accelerated deactivation of its substrate, $G\alpha_o$ (Fig. 2, B and D). Application of glycine had no effect on either baseline $G\alpha_o$ deactivation or the RGS7-Gβ5–assisted process (Fig. 2, B and D). However, when GPR158 was coexpressed together with RGS7-Gβ5, glycine significantly decelerated $G\alpha_o$ deactivation (Fig. 2, C and D), suggesting that it specifically inhibited the GAP activity of RGS7-Gβ5 by engaging GPR158. Dose–response studies showed that the median inhibitory concentration (IC_{50}) of glycine on GPR158 is ~3 μM (Fig. 2E). Taurine displayed a similar inhibitory effect on $G\alpha_o$ deactivation only in cells coexpressing GPR158

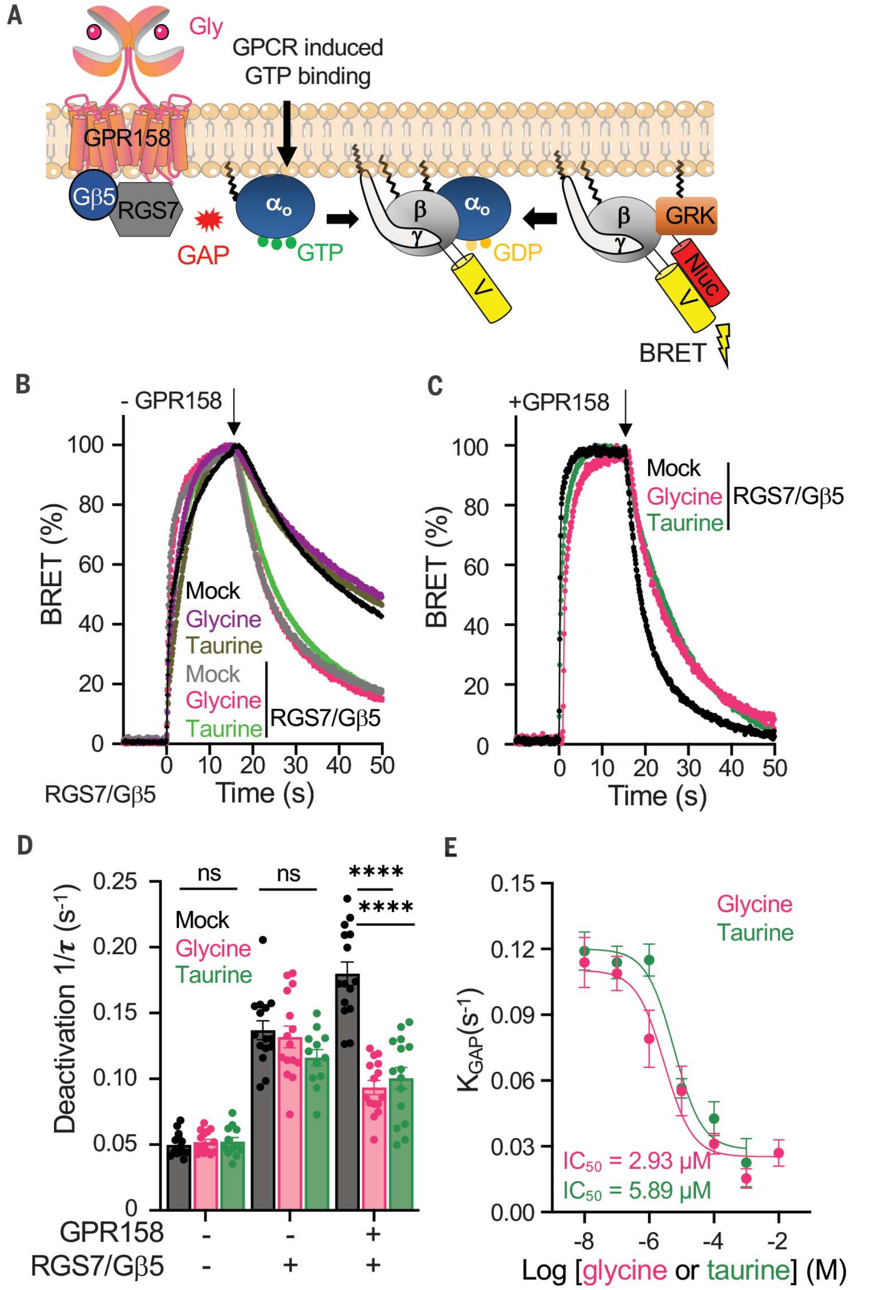


Fig. 2. Glycine and taurine slow deactivation of $G\alpha_o$ by GPR158-RGS7-Gβ5 complex. (**A**) Schematics of the BRET-based GAP assay. G proteins are activated at t = 0 s by stimulating GPCR (dopamine DR receptor, 0.1 mM). After reaching steady state, the GPCR activity is terminated by injection of haloperidol (0.1 mM) at t = 15 s (arrow). G protein deactivation is then monitored by following quenching of the BRET signal. (**B** and **C**) Traces of BRET signal showing $G\alpha_o$ activation and deactivation time course with or without glycine or taurine (100 μM) treatment in cells without GPR158 (B) or cells transfected with GPR158 (C). (**D**) Quantification of deactivation time constant of the reactions presented in (B) and (C). $1/\tau$ is calculated from deactivation curves of n = 5 independent experiments conducted in triplicate from each cell transfection group. Data represent mean ± SEM. ****$p < 0.0001$, ns (not significant) = $p > 0.05$, two-way ANOVA. (**E**) Dose–response profile of changes in GAP activity (K_{GAP}) calculated by subtracting the baseline deactivation rate ($1/\tau$) from the rate of the reaction in the presence of GPR158-RGS7-Gβ5. Data represent mean ± SEM of n = 4 independent experiments conducted in triplicate.

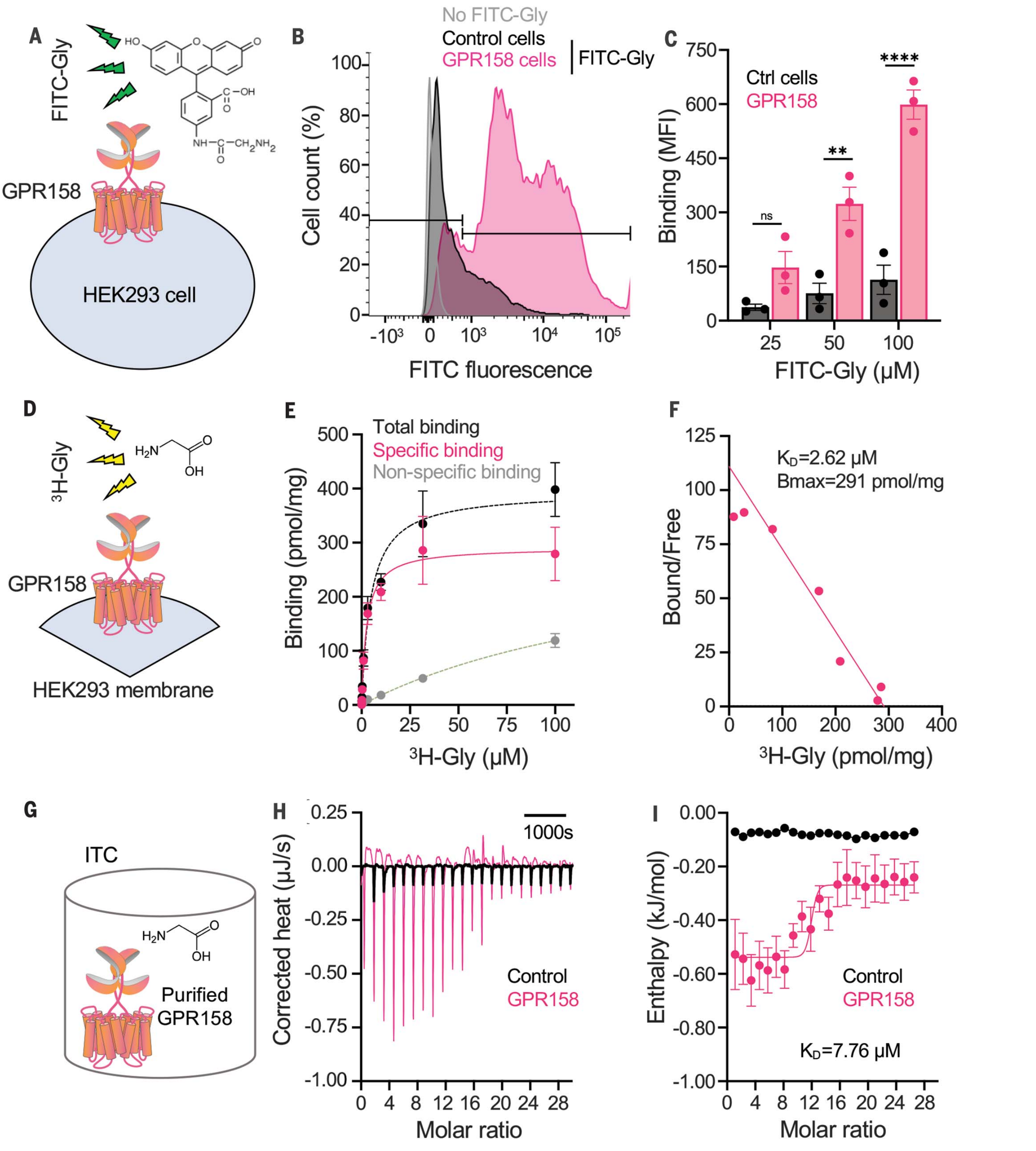


Fig. 3. Direct interaction of glycine with GPR158. (**A**) Schematics of assay design for detecting glycine binding to GPR158 by flow cytometry. (**B**) Flow cytometry histogram showing distribution of cellular populations after sorting. (**C**) Quantification of FITC-glycine binding detected in flow cytometry experiments. The median of fluorescence (MFI) is quantified and plotted. Error bars indicate SEM, $n = 3$, $^{**}p < 0.01$, $^{****}p < 0.0001$, two-way ANOVA. (**D**) Schematics of the radioligand binding assay. (**E**) Quantification of [^{3}H]glycine binding to membrane expressing GPR158. Data show mean of four independent experiments; error bars indicate SEM, $n = 4$. (**F**) Scatchard plot of the [^{3}H]glycine radioligand binding assay. Data show mean of four independent experiments; error bars indicate SEM, $n = 4$. (**G**) Schematics of the assay design detecting glycine binding to GPR158 by isothermal titration calorimetry (ITC) with purified protein. (**H**) ITC binding profile showing glycine binding to GPR158 in the initial run with fresh sample. (**I**) Quantification of binding determined by fitting the integrated isotherm to an independent binding model. Data show mean of four experimental runs; error bars indicate SEM.

with RGS7-Gβ5, but with a lower IC_{50} of ~6 μM (Fig. 2, B to E).

We further tested whether glycine or taurine could induce GPR158 to activate G proteins as canonical GPCRs do (fig. S2A). We observed no significant activation of any G proteins tested with either glycine or taurine (fig. S2, B to I). We also tested whether glycine could induce β-arrestin recruitment to GPR158 using a BRET assay and obtained no significant response (fig. S3).

GPR158 directly binds glycine

To confirm that GPR158 is a direct target of glycine, we used several strategies. First, we devised a flow cytometry–based assay to monitor binding of fluorescein isothiocyanate (FITC)–

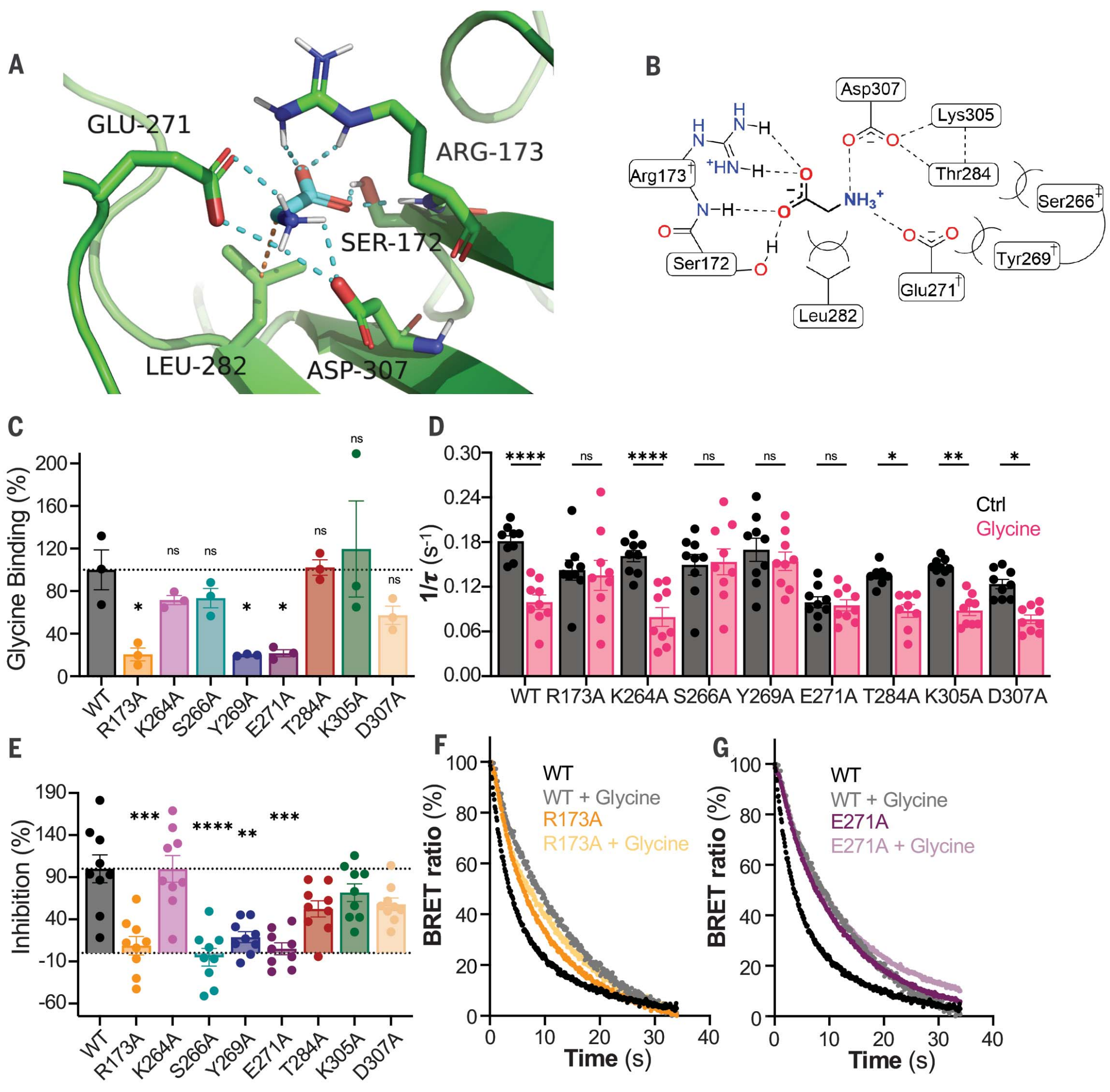


Fig. 4. Probing Cache domain of GPR158 as a ligand binding site. (**A**) Computational docking of glycine (teal) into putative ligand-binding pocket on GPR158 Cache domain (green). Glycine and directly interacting residues are shown as sticks. Hydrogen bonds (teal) and van der Waals interactions (orange) are shown as dotted lines. (**B**) Diagram showing interactions in the docked model of glycine against GPR158 Cache domain. Hydrogen bonds are shown as dashed lines, van der Waals interactions are shown as intersecting semicircles, and the secondary structural context of Ser266 and Tyr269 is shown as an arc. Dagger indicates a residue implicated in glycine binding in (A). (**C**) Radioligand binding assay of [^{3}H]glycine in cells expressing GPR158 mutants. Data show mean ± SEM of three independent experiments, *$p < 0.05$, ns = $p > 0.05$, one-way ANOVA. (**D**) Functional evaluation of GPR158 mutants in GAP BRET assay. Error bars indicate mean ± SEM of three independent experiments conducted in triplicate. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, ns = $p > 0.05$, two-way ANOVA. (**E**) Quantification of glycine inhibitory effect on K_{GAP} normalized to the effect seen with wild-type (WT) receptor. Error bars indicate mean ± SEM of three independent experiments conducted in triplicate. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, ns = $p > 0.05$, one-way ANOVA. (**F** and **G**) Traces of $G\alpha_o$ deactivation time course upon glycine addition. Single-letter abbreviations for the amino acid residues are as follows: D, Asp; E, Glu; K, Lys; R, Arg; S, Ser; T, Thr; and Y, Tyr.

conjugated glycine to cells expressing GPR158 (Fig. 3A). When HEK293 cells expressing GPR158 were incubated with FITC-glycine, we observed labeling of a significant population of cells (Fig. 3B). No such labeling was evident when FITC-glycine was incubated with cells not transfected with GPR158. Dose–response studies further confirmed this binding and its selectivity across the ranges of glycine used (Fig. 3C).

Next, we performed radioligand binding assays examining binding of [^{3}H]-labeled glycine to HEK293 cells expressing GPR158 (fig. S4A). We detected significant binding of [^{3}H]glycine to GPR158-expressing cells across concentrations (Fig. S4B). We isolated cellular membranes and conducted classical radioligand titration experiments (Fig. 3D). We detected saturable [^{3}H]glycine binding to membranes containing GPR158 in substantial excess over linear nonspecific binding to membranes devoid of GPR158 (Fig. 3E). Scatchard analysis (Fig. 3F) estimated the dissociation constant, K_D, of GPR158 for glycine to be ~3 μM. Binding competition experiments directly comparing the ability of glycine and taurine to displace [^{3}H]glycine bound to GPR158 (fig. S5) confirmed the specificity of glycine and taurine binding to GPR158 and also revealed a twofold lower affinity of taurine relative to glycine (IC_{50}: ~3 μM versus ~6 μM).

In addition, we examined binding of nonlabeled glycine directly to purified GPR158 using isothermal titration calorimetry (ITC) (Fig. 3G). Titration experiments showed saturation of the heat released upon glycine addition to GPR158 yielding K_D ranging from 2 to 16 μM across experiments conducted first with a fresh sample (Fig. 3, H and I) and subsequently rerun after removal of glycine and detergent (fig. S6). The affinity of glycine obtained in direct binding experiments is in good agreement with the affinity measured in the functional GAP assays, indicating that binding to glycine is responsible for changes in GPR158 activity.

Glycine binds to Cache domain of GPR158 and modulates GAP activity of RGS7-Gβ5 complex

We performed molecular docking experiments fitting glycine into a model of the putative ligand-binding pocket in the Cache domain of GPR158, built by supplementing experimental structure (*30*) with a missing loop taken from the AlphaFold2 prediction (Fig. 4A and table S1). Although another structure of GPR158 is available (*31*), it did not resolve side-chain conformations and thus was not considered

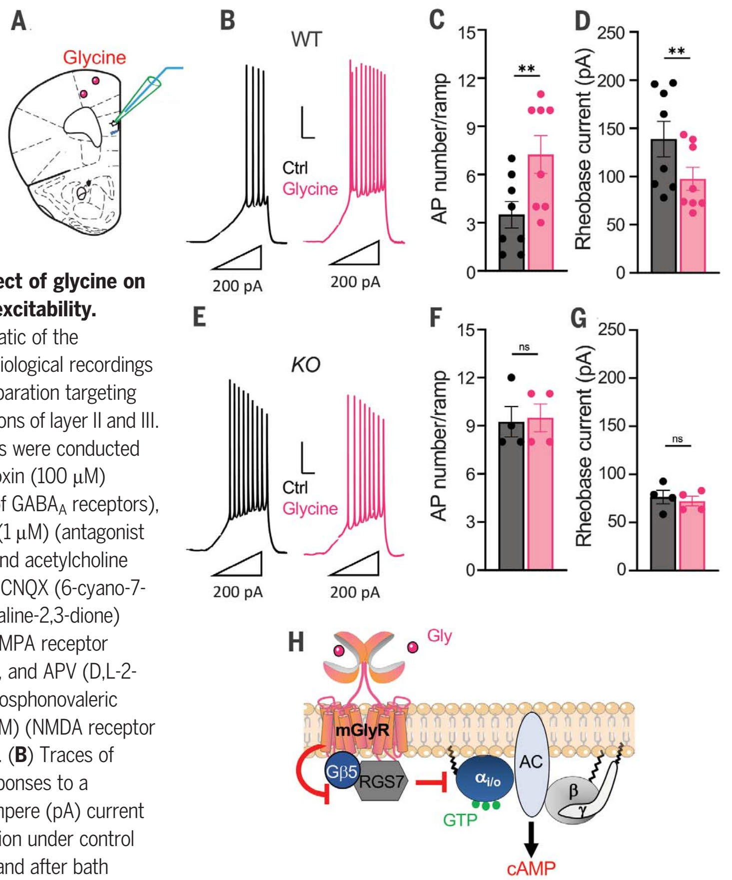


Fig. 5. Effect of glycine on neuronal excitability. (**A**) Schematic of the electrophysiological recordings in slice preparation targeting mPFC neurons of layer II and III. Experiments were conducted with picrotoxin (100 μM) (blockade of $GABA_A$ receptors), strychnine (1 μM) (antagonist of glycine and acetylcholine receptors), CNQX (6-cyano-7-nitroquinoxaline-2,3-dione) (20 μM) (AMPA receptor antagonist), and APV (D,L-2-amino-5-phosphonovaleric acid) (50 μM) (NMDA receptor antagonist). (**B**) Traces of voltage responses to a 200-picoampere (pA) current ramp injection under control conditions and after bath application of glycine (1 mM). (**C**) Quantification of changes in excitability by number of action potentials fired in response to 200-pA current ramp (*n* = eight neurons from five mice). (**D**) Quantification of changes in excitability by rheobase current under control condition and glycine application (*n* = eight neurons from five mice). (**E**) Traces of voltage responses to a 200-pA current ramp injection obtained from layer II and III pyramidal neurons in *Gpr158* KO mice under control conditions and during bath application of glycine (1 mM). (**F**) Quantification of changes in excitability by number of action potentials (AP) fired in response to 200-pA current ramp (*n* = four neurons from three mice). (**G**) Quantification of changes in excitability by rheobase current under control condition and glycine application (*n* = four neurons from three mice). (**H**) Schematic representation of the proposed mechanism of glycine effects on mGlyR. In all graphs, nonparametric *t* test; Wilcoxon test was used for statistical analysis, ns = $p > 0.05$, $^{**}p < 0.01$.

as a source of alternate receptor conformations for docking. For the best-scored glycine pose, glycine could be well accommodated in a pocket where it is stabilized by a network of hydrogen-bonding interactions with S172, R173, E271, and D307 side chains, with the charged side chains ideally positioned to stabilize the carboxylate and amine moieties of the zwitterion. These residues are embedded in a web of other hydrophilic residues located in a close vicinity (e.g., K264, S266, Y269, T284, and K305) lining the pocket (Fig. 4B). Docking studies performed with taurine found a cluster of poses that overall matched the putative binding mode of glycine, retaining the features described for glycine (fig. S7). The size of the pocket is spatially constrained, particularly by L282, in such a way that other amino acids cannot be easily accommodated without steric clashes with side chains of residues lining the pocket, which provides a possible explanation for the selectivity of the recognition (fig. S8).

To test the role of the residues forming the putative glycine pocket in the GPR158 Cache domain, we performed site-directed mutagenesis. In radioligand binding assays, the R173A, E271A, and Y269A mutants showed near complete loss of [^{3}H]glycine binding, confirming the essential role of these residues in ligand coordination (Fig. 4B). We then tested each of the mutants in functional assays (Fig. 4, C and D). Each of the mutants defective in glycine binding also lost an ability to inhibit the GAP activity of RGS7-Gβ5 (Fig. 4E). The activity of the S266A mutant, which normally binds glycine, was not regulated by it, suggesting that some of the residues in the binding pocket are involved in conformational transitions triggered by ligand interaction (*37*). The mechanism by which mutating E271 residue resulted in loss of glycine responsiveness also deviated for that of other mutations. This mutant exhibited a much slower deactivation kinetics in the absence of glycine, generating a constitutively inhibited receptor.

Glycine modulates neuronal excitability through GPR158

Lastly, we assessed the impact of glycine modulation of GPR158 on neuronal activity. We examined the intrinsic properties of layer II and III neurons in the prelimbic cortex, where GPR158 is prominently expressed (*25*) and regulates neuronal excitability (*27*). The metabotropic effects of glycine are not well characterized across the nervous system. Therefore, we started by defining the effects of glycine on layer II and III neurons. To isolate metabotropic actions, we antagonized excitatory and inhibitory synaptic drive with pharmacological blockade and measured the current-voltage relation in response to a depolarizing current ramp. Application of glycine significantly increased the number of action potentials while decreasing the amount of current necessary to elicit the first action potential (Fig. 5, A to C) without changes in the resting membrane potential (fig. S9). This excitatory effect of glycine is distinct from its canonical inhibitory action mediated by glycine receptor (GlyR) ion channels. Interestingly, glycine application did not produce any changes in the intrinsic excitability of layer V neurons (fig. S9), which do not express GPR158 (*27*).

To confirm the involvement of GPR158 in the effects of glycine, we studied *Gpr158* knockout (*Gpr158* KO) mice. Glycine application failed to alter excitability of layer II and III neurons in prefrontal cortex of *Gpr158* KO mice (Fig. 5, D to F). We also tested the effect of taurine on the excitability of layer II and III neurons (fig. S11). These experiments revealed small effects on neuronal firing in the same direction as glycine. However, these effects did not reach the criteria for statistical significance, possibly because of the lower efficacy of taurine.

Discussion

In this study, we demonstrated that GPR158 serves as a metabotropic receptor for glycine. We also report that GPR158 can be modulated by taurine, which acts as a partial agonist for this receptor. This finding was enabled by recently obtained high-resolution structure of the receptor, which revealed the presence of a ligand-binding module: the Cache domain. Cache domains are well-known receptors for amino acids and other related small molecules ubiquitously used by bacterial chemoreceptors. Only two GPCRs contain them, including GPR158 and night-blindness associated receptor GPR179 (*38*), whose ligand remains to be established. We present evidence that glycine acts as a bona fide ligand on GPR158,

including direct binding and resultant change in receptor activity eliciting cellular response. This puts GPR158 in line with other class C GPCRs, many of which are amino acid sensors, such as the metabotropic glutamate receptors (mGluRs) and the receptor for γ-aminobutyric acid (GABA), $GABA_B$. Thus, we propose a generic name for GPR158 to be metabotropic glycine receptor, or mGlyR. We did not observe significant GPR158-mediated neuronal responses to taurine in cortical neurons, consistent with weaker effects of taurine on GPR158 relative to glycine. However, it remains possible that GPR158 may still mediate the effects of taurine in other neuronal populations or under certain conditions, possibly making GPR158 a receptor for both glycine and taurine.

The mechanism by which mGlyR (GPR158) signals upon glycine or taurine binding deviates from canonical actions of GPCRs. Instead of activating G proteins, mGlyR recruits a RGS7-Gβ5 complex, docking it into the intracellular pocket that canonical GPCRs use for interacting with G proteins and relaying changes in seven-transmembrane architecture upon ligand binding into conformational changes in Gα, triggering nucleotide exchange. Thus, in the model we propose (Fig. 5H), glycine binding to the Cache domain of mGlyR changes the conformation of the intracellular surface, which in turn affects conformation of RGS7-Gβ5. This change reduces the ability of RGS7-Gβ5 to stimulate Gα GTPase, likely by disfavoring its orientation toward the membrane. In this sense, glycine serves as an antagonist of the GPR158-RGS7-Gβ5 complex by reducing its activity. Because RGS7-Gβ5 is a selective GAP for the inhibitory $G_{i/o}$ proteins, which regulate cAMP production (*39*, *40*), inhibition of RGS7-Gβ5 activity via GPR158 influences cAMP levels. The direction of the effect on the cAMP production is likely determined by the identity of the adenylyl cyclases present in a particular cell, as they are known to be differentially regulated by $G\alpha_i$ and $G\alpha_o$ (via Gβγ) (*41*). Thus, glycine signals via mGlyR by inhibiting inhibitory G protein regulation, thereby generating an excitatory influence. This regulation endows the metabotropic glycinergic system with a distinct feature that makes the degree of its influence scale with the extent of $G_{i/o}$ activation by other GPCR cascades, with its influence increasing upon the increase in $G_{i/o}$ inputs.

The discovery of mGlyR also opens many interesting avenues for exploring the metabotropic influence of glycine and its role in nervous system physiology. Indeed, metabotropic effects of glycine have been anecdotally noted (*6*, *7*, *42*), but molecular and circuit dissection of this influence have been limited. The relatively high affinity of mGlyR for glycine (~3 μM) should allow it to signal without concomitant engagement of GlyRs, which have an order-of-magnitude-lower affinity for glycine, creating an independent neuromodulatory channel (*6*). The mGlyR effects on neurons that we observe are also excitatory, contrasting with the largely inhibitory influence of ionotropic GlyR receptors (*9*, *43*). The two systems likely overlap and are involved in autotuning and homeostatic feedback, as has been noted for other pairs of ionotropic and metabotropic systems. Thus, in the context of intact neural circuitry, glycine likely triggers more complex responses that may involve interplay between ionotropic and metabotropic systems.

Furthermore, we think that glycinergic signaling by means of mGlyR has implications for understanding mood disorders and for the development of new pharmacological strategies. mGlyR is prominently expressed in the medial prefrontal cortex (mPFC) (*25*), a region critically involved in depression (*44*). Glycine and its transporters are also colocalized in the mPFC (*45*–*47*). Both taurine and glycine have been heavily implicated in the pathophysiology of depression (*8*–*10*, *48*) and are dysregulated in plasma of humans diagnosed with major depressive disorder (*10*). Furthermore, taurine has an antidepressant effect on stress-induced depressive rats (*49*). Because these amino acids inhibit mGlyR, and because knockout of mGlyR in mice also results in an antidepressive phenotype and stress resilience (*25*), it seems possible that antidepressant properties of glycine and taurine may be mediated by mGlyR. The ubiquitous nature and multitude of the effects limit the potential of glycine and taurine to be used as medications. However, identification of mGlyR presents a new target for the development of antidepressants that we postulate to be small molecules that selectively inhibit this receptor to avoid possibly related receptors, such as GPR179 in the eye.

REFERENCES AND NOTES

1. M. S. Hernandes, L. R. P. Troncone, *J. Neural Transm.* **116**, 1551–1560 (2009).
2. P. Legendre, *Cell. Mol. Life Sci.* **58**, 760–793 (2001).
3. L. G. Harsing Jr., P. Matyus, *Brain Res. Bull.* **93**, 110–119 (2013).
4. H. Zhu, E. Gouaux, *Nature* **599**, 513–517 (2021).
5. I. Pérez-Torres, A. M. Zuniga-Munoz, V. Guarner-Lans, *Mini Rev. Med. Chem.* **17**, 15–32 (2017).
6. Y. Han, J. Zhang, M. M. Slaughter, *J. Neurosci.* **17**, 3392–3400 (1997).
7. M. Hou, L. Duan, M. M. Slaughter, *J. Physiol.* **586**, 2913–2926 (2008).
8. C. C. Huang *et al.*, *Biol. Psychiatry* **74**, 734–741 (2013).
9. W. Li *et al.*, *Neuropharmacology* **157**, 107688 (2019).
10. C. Altamura, M. Maes, J. Dai, H. Y. Meltzer, *Eur. Neuropsychopharmacol.* **5** (suppl. 1), 71–75 (1995).
11. N. Wettschureck, S. Offermanns, *Physiol. Rev.* **85**, 1159–1204 (2005).
12. E. J. Neer, *Cell* **80**, 249–257 (1995).
13. D. Yang *et al.*, *Signal Transduct. Target. Ther.* **6**, 7 (2021).
14. E. V. Gurevich, V. V. Gurevich, *Genome Biol.* **7**, 236 (2006).
15. A. K. Shukla, K. Xiao, R. J. Lefkowitz, *Trends Biochem. Sci.* **36**, 457–469 (2011).
16. E. Hermans, *Pharmacol. Ther.* **99**, 25–44 (2003).
17. I. Masuho *et al.*, *Cell* **183**, 503–521.e19 (2020).
18. M. Abramow-Newerly, A. A. Roy, C. Nunn, P. Chidiac, *Cell. Signal.* **18**, 579–591 (2006).
19. A. Fajardo-Serrano *et al.*, *Hippocampus* **23**, 1231–1245 (2013).
20. J. Celver, M. Sharma, A. Kovoor, *J. Neurochem.* **120**, 56–69 (2012).
21. K. Psifogeorgou *et al.*, *J. Neurosci.* **31**, 5617–5624 (2011).
22. C. Orlandi *et al.*, *J. Cell Biol.* **197**, 711–719 (2012).
23. J. A. Stockert, L. A. Devi, *Front. Pharmacol.* **6**, 100 (2015).
24. C. Laschet, N. Dupuis, J. Hanson, *Biochem. Pharmacol.* **153**, 62–74 (2018).
25. L. P. Sutton *et al.*, *eLife* **7**, e33273 (2018).
26. C. Orlandi, L. P. Sutton, B. S. Muntean, C. Song, K. A. Martemyanov, *Neuropsychopharmacology* **44**, 642–653 (2019).
27. C. Song, C. Orlandi, L. P. Sutton, K. A. Martemyanov, *J. Biol. Chem.* **294**, 13145–13157 (2019).
28. D. Çetereisi *et al.*, *Front. Cell. Neurosci.* **13**, 465 (2019).
29. L. Khrimian *et al.*, *J. Exp. Med.* **214**, 2859–2873 (2017).
30. D. N. Patil *et al.*, *Science* **375**, 86–91 (2022).
31. E. Jeong, Y. Kim, J. Jeong, Y. Cho, *Nat. Commun.* **12**, 6805 (2021).
32. L. I. Jiang *et al.*, *J. Biol. Chem.* **282**, 10576–10584 (2007).
33. O. Kletke, G. Gisselmann, A. May, H. Hatt, O. A. Sergeeva, *PLOS ONE* **8**, e61733 (2013).
34. N. Hussy, C. Deleuze, A. Pantaloni, M. G. Desarménien, F. Moos, *J. Physiol.* **502**, 609–621 (1997).
35. I. Masuho, K. Xie, K. A. Martemyanov, *J. Biol. Chem.* **288**, 25129–25142 (2013).
36. I. Masuho, K. A. Martemyanov, N. A. Lambert, *G Protein-Coupled Receptors in Drug Discovery: Methods and Protocols*, M. Filizola, Ed. (Springer, 2015), pp. 107–113.
37. B. W. H. Liauw, H. S. Afsari, R. Vafabakhsh, *Nat. Chem. Biol.* **17**, 291–297 (2021).
38. T. A. Ray *et al.*, *J. Neurosci.* **34**, 6334–6343 (2014).
39. B. S. Muntean *et al.*, *Cell Rep.* **34**, 108718 (2021).
40. P. Sassone-Corsi, *Cold Spring Harb. Perspect. Biol.* **4**, a011148–a011148 (2012).
41. R. Sadana, C. W. Dessauer, *Neurosignals* **17**, 5–22 (2009).
42. E. Albiñana, S. Sacristán, R. Martín del Río, J. M. Solís, J. M. Hernández-Guijo, *Cell. Mol. Neurobiol.* **30**, 1225–1233 (2010).
43. M. H. Aprison, N. S. Nadi, *Amino Acids as Chemical Transmitters*, F. Fonnum, Ed. (Springer US, 1978), pp. 531–570.
44. B. D. Hare, R. S. Duman, *Mol. Psychiatry* **25**, 2742–2758 (2020).
45. H. U. Zeilhofer *et al.*, *J. Comp. Neurol.* **482**, 123–141 (2005).
46. N. Kawai *et al.*, *Amino Acids* **42**, 2129–2137 (2012).
47. R. J. Harvey, B. K. Yee, *Nat. Rev. Drug Discov.* **12**, 866–885 (2013).
48. S. Kim, K. B. Hong, S. Kim, H. J. Suh, K. Jo, *Sci. Rep.* **10**, 11370 (2020).
49. G. F. Wu *et al.*, *Sci. Rep.* **7**, 4989 (2017).
50. T. Laboute *et al.*, Orphan receptor GPR158 serves as a metabotropic glycine receptor: mGlyR, Zenodo (2023); https://doi.org/10.5281/zenodo.7631532.

ACKNOWLEDGMENTS

We thank N. Martemyanova for help with mouse husbandry, X. Li for experimental help, and members of the Martemyanov Lab for helpful discussions. We thank Servier Medical Art for templates used in graphics. **Funding:** This work was supported by NIH grants MH105482 (to K.A.M.) and GM069832 (to S.F.). **Author contributions:** T.L. and K.A.M. conceived the project; T.L. performed all functional experiments; S.Z. performed electrophysiology experiments; C.G., D.P., S.F., and M.H. performed and analyzed computational docking and structural modeling; B.A.W. and R.N.R. performed and analyzed ITC experiments; T.L. and K.A.M. wrote the manuscript with input from all other authors; and K.A.M. supervised the project. **Competing interests:** T.L. and K.A.M. are listed as inventors on a patent application describing methods to study GPR158 activity; K.A.M. is a cofounder of Blueshield Therapeutics, a startup company pursuing GPR158 as a drug target. **Data and materials availability:** All data that support the findings are available at Zenodo (*50*). All plasmids generated during this study are freely available upon request.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add7150
Materials and Methods
Figs. S1 to S10
Table S1
References (*51*–*65*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 29 June 2022; resubmitted 4 November 2022
Accepted 3 March 2023
10.1126/science.add7150

Check for updates

FUNCTION PREDICTION

Enzyme function prediction using contrastive learning

Tianhao Yu[1,2,3]†, Haiyang Cui[1,2,3]†, Jianan Canal Li[3,4], Yunan Luo[5], Guangde Jiang[1,2], Huimin Zhao[1,2,3,6]*

Enzyme function annotation is a fundamental challenge, and numerous computational tools have been developed. However, most of these tools cannot accurately predict functional annotations, such as enzyme commission (EC) number, for less-studied proteins or those with previously uncharacterized functions or multiple activities. We present a machine learning algorithm named CLEAN (contrastive learning–enabled enzyme annotation) to assign EC numbers to enzymes with better accuracy, reliability, and sensitivity compared with the state-of-the-art tool BLASTp. The contrastive learning framework empowers CLEAN to confidently (i) annotate understudied enzymes, (ii) correct mislabeled enzymes, and (iii) identify promiscuous enzymes with two or more EC numbers—functions that we demonstrate by systematic in silico and in vitro experiments. We anticipate that this tool will be widely used for predicting the functions of uncharacterized enzymes, thereby advancing many fields, such as genomics, synthetic biology, and biocatalysis.

The development of DNA sequencing technologies, and particularly genomics and metagenomics tools, has led to the discovery of numerous protein sequences from organisms across all branches of life. For example, UniProt Knowledgebase has cataloged ~190 million protein sequences. However, only <0.3% (approximately half a million) of these proteins were reviewed by human curators, out of which <19.4% are supported by clear experimental evidence (*1*). Consequently, protein function annotation is highly dependent on computational annotation methods. However, the study on large-scale, community-based critical assessment of protein function annotation (CAFA) found that ~40% of the automatically annotated enzymes using existing computational tools are incorrectly annotated (*2*). Therefore, functional annotation of proteins remains an overwhelming challenge in protein science. Particularly, the inequality in protein annotation of understudied and promiscuous proteins has impeded biomedical progress and drug discovery (*3*, *4*).

Enzyme commission (EC) number is the most well-known numerical classification scheme of enzymes, which specifies the catalytic function of an enzyme by four digits. Because experimental characterization of the function of a target enzyme is often laborious and expensive, numerous computational tools for enzyme function annotation have been developed (*1*, *5*, *6*). They include but are not limited to sequence similarity–based (*7*–*9*), homology-based (*10*, *11*), structure-based (*12*, *13*), and machine learning (ML)–based (*14*, *15*) approaches. Among them, sequence similarity–based Basic Local Alignment Search Tools for proteins (BLASTp) is the most widely used tool (*7*). However, BLASTp and other alignment tools annotate functions based solely on sequence similarity, making the prediction result less reliable when sequence similarity is low. On the other hand, almost all the existing ML models, such as DeepEC (*14*) and ProteInfer (*15*), are based on a multilabel classification framework and suffer from the limited and imbalanced training dataset that is common in biology. Therefore, a robust tool with better accuracy and EC coverage is required to unlock the potential of currently uncharacterized proteins and to understand the range of protein functions.

In this work, we report a ML model named CLEAN (contrastive learning–enabled enzyme annotation) for enzyme function prediction. CLEAN was trained on high-quality data from UniProt, taking amino acid sequence as input and outputting a list of enzyme functions (EC numbers as the example) ranked by the likelihood. To validate the accuracy and robustness of CLEAN, we performed extensive in silico experiments. Furthermore, we challenged CLEAN to annotate EC numbers for an in-house collected database of all uncharacterized halogenases (36 in total) followed by case studies as in vitro experimental validation. CLEAN outperformed other EC number annotation tools at these tasks, including BLASTp and state-of-the-art ML models.

Model development and evaluation

Unlike previously developed ML algorithms that frame EC number prediction tasks as a multilabel classification problem, CLEAN used a contrastive learning (*16*, *17*) framework. Our training objective is to learn an embedding space of enzymes where the Euclidean distance reflects the functional similarities. The embedding refers to a numerical representation (vectors or matrices) of protein sequences that is readable by machine while still retaining the important features and information carried by the enzyme. In CLEAN's task, the amino acid sequences with the same EC number have a small Euclidean distance, whereas sequences with different EC numbers have a large distance. Contrastive losses were used to train the model with supervision (*16*, *18*). During the training process (Fig. 1A), each reference sequence (anchor) in the training dataset was sampled with a sequence with the same EC number (positive) and a sequence with a different EC number (negative). Aiming to facilitate training efficiency by providing the model with challenging negative samples—instead of drawing them randomly—negative sequences with embeddings that had a small Euclidean distance with the anchor were prioritized.

In the training stage, the protein representation obtained from the language model ESM-1b (*19*) was used as the input of a feedforward neural network, whose output layer produced a refined, function-aware embedding of the input protein. The learning objective is a contrastive loss function that minimizes the distance between the anchor and the positive while maximizing the distance between the anchor and the negative. When making predictions, the representation of an EC number cluster center was obtained by averaging the learned embeddings of all sequences in the training set belonging to that EC number (Fig. 1B). Subsequently, the pairwise distances between the query sequence and all EC number cluster centers were calculated. EC numbers of clusters that are significantly close to the query sequence are predicted as the EC numbers for the input protein (supplementary text, section 1).

The database used for model development and evaluation was a universal protein knowledgebase UniProt (*1*). Two EC selection methods were developed to predict confident EC numbers from the output ranking (Fig. 1C): (i) a greedy approach that selects EC numbers that have the maximum separation (stand out) from other EC numbers in terms of the pairwise distance to the query sequence and (ii) a *P* value–based method that identifies EC numbers with statistical significance compared with background (see materials and methods). On a train-test split in which none of the enzymes in the test set share >50% identity with any enzymes in the training set, using the maximum-separation selection method, CLEAN achieved a 0.865 F1 score—a commonly used accuracy metric indicating the harmonic mean of precision and recall. Even at 10% sequence identity clustering, CLEAN reached a 0.67 F1 score. Additionally, CLEAN achieved much higher performance compared with the baseline method using ESM-1b without contrastive learning (fig. S1).

[1]Department of Chemical and Biomolecular Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. [2]Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. [3]National Science Foundation Molecule Maker Lab Institute, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. [4]Department of Computer Science, Cornell University, Ithaca, NY 14850, USA. [5]School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30308, USA. [6]US Department of Energy Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA.
***Corresponding author. Email: zhao5@illinois.edu**
†These authors contributed equally to this work.

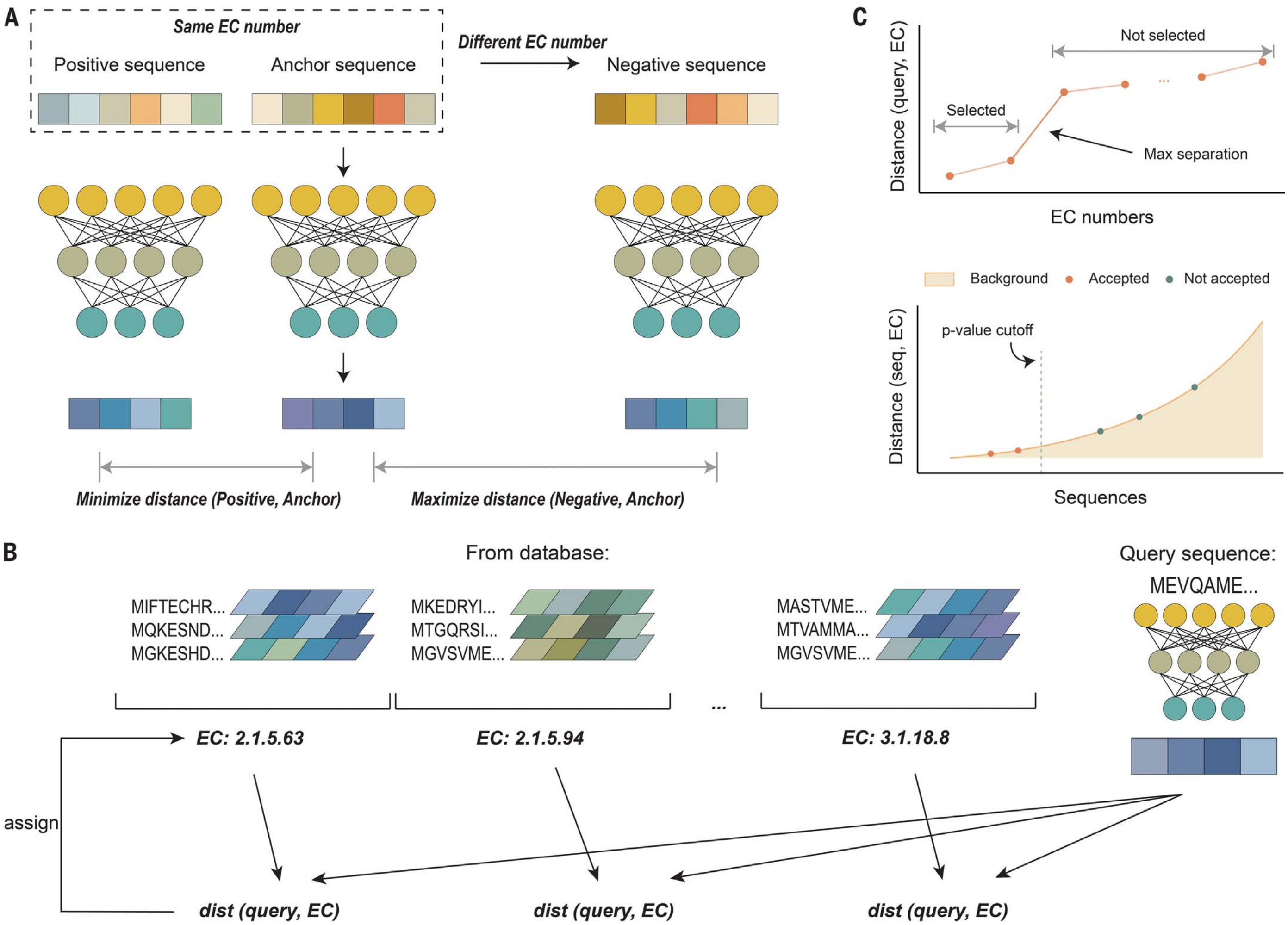


Fig. 1. The contrastive learning–based framework of CLEAN. (**A**) During training, positives and negatives were sampled on the basis of EC numbers. The input sequences were embedded and passed through a neural network. The series of squares with warm colors stands for the representation of input sequence embedded by ESM-1b. Similarly, the sequence embeddings obtained from the supervised contrastive learning neural network are illustrated by cool colors. (**B**) The representations of an EC number are obtained by averaging the representations of enzymes under this EC number. When predicting the EC number, the query sequence embedding was compared with each EC number's representation (shown as a parallelogram with cool colors) to obtain the pairwise Euclidean distance between the query sequence and each EC number. The distance reflects the similarity between EC numbers and the query sequence. (**C**) When used as a classification model, two methods, maximum separation (above) and *P* value (below), were implemented to prioritize confident predictions of EC numbers from the ranking order.

Benchmarking CLEAN with previous EC number annotation tools

After training, the prediction performance of CLEAN was systematically investigated by comparing it with six state-of-the-art EC number annotation tools [i.e., ProteInfer (*15*), DeepEC (*14*), BLASTp, DEEPre (*20*), CatFam (*21*), and ECPred (*22*)]. Two independent datasets not included in any model's development were used to deliver a fair and rigorous benchmark study. The first dataset, New-392, consisted of 392 enzyme sequences covering 177 different EC numbers, containing data from Swiss-Prot released after CLEAN was trained (April 2022). The prediction scenario represented a practical situation, where the labeled knowledgebase was the Swiss-Prot database and functions of query sequences were unknown. Overall, CLEAN resulted in the highest value in various multilabel accuracy metrics, including precision (0.597) and recall (0.481), when compared with ProteInfer and DeepEC (Fig. 2A). Also, CLEAN achieved an F1 score of 0.499, whereas ProteInfer and DeepEC had scores of 0.309 and 0.230, respectively.

The second independent dataset, denoted as Price-149, was a set of experimentally validated results described by Price *et al.* (*23*). The Price-149 dataset was first curated by ProteInfer (*15*) as a challenging dataset because the existing sequences were determined to be incorrectly or inconsistently labeled in databases like Kyoto Encyclopedia of Genes and Genomes (KEGG) by automated annotation methods. Again, CLEAN achieved the highest F1 score (0.495) compared with BLASTp, ProteInfer, and DeepEC (Fig. 2B). Notably, in this challenging task, CLEAN had a 3.0-fold higher F1 score than ProteInfer (0.166) and an almost 5.8-fold higher score than DeepEC (0.085). The evaluations on the New-392 and Price-149 datasets demonstrate that CLEAN is more precise and reliable than previously developed ML-based models for predicting functions for newly discovered proteins, especially the ones without known enzyme functions.

Understanding CLEAN's performance on annotating understudied EC number

Next, we investigated why CLEAN performs better than other ML models on understudied EC numbers. We curated a validation dataset with enzymes from rare EC numbers to test our hypothesis that, compared with the multilabel

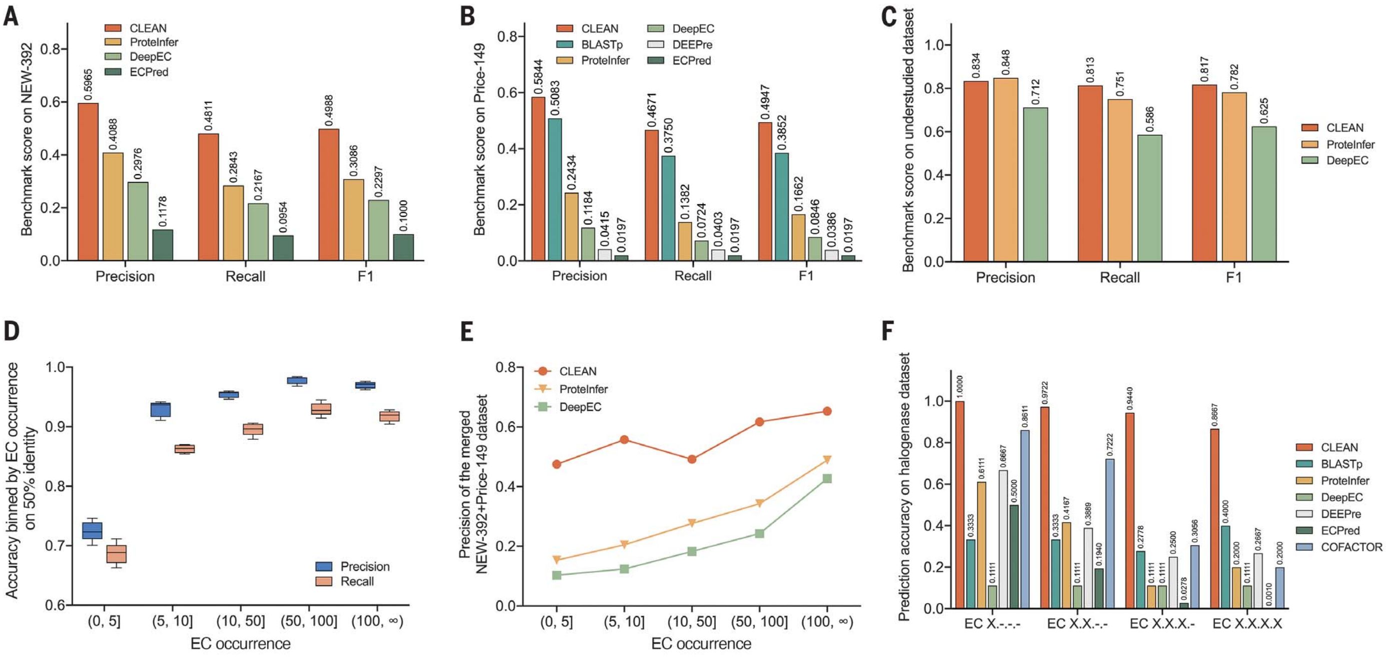


Fig. 2. Quantitative comparison of CLEAN with the state-of-the-art EC number prediction tools. (**A**) Evaluation of CLEAN's performance toward three multilabel accuracy metrics (precision, recall, and F1 score) examined on the New-392 database. Four top-ranked models, ProteInfer, DeepEC, CatFam, and ECPred, were used for comparison. (**B**) Comparison of CLEAN, BLASTp, ProteInfer, DeepEC, DEEPre, CatFam, and ECPred on the Price-149 database. (**C**) Comparison of CLEAN, ProteInfer, and DeepEC on a dataset of underrepresented EC numbers. (**D**) The accuracy binned plot of CLEAN using the test set with <50% identity to the training set evaluated with SupconH loss. Precision and recall values were binned by the number of times that the EC number appeared in the training set—i.e., the bin (0,5] means that the EC numbers occurs less than five times in the training set. The box plots show the results of fivefold cross-validation. (**E**) Evaluation on the combined datasets of Price-149 and New-392 binned by the number of times that the EC number appeared in CLEAN's training dataset. (**F**) Prediction accuracy of CLEAN on an in-house–curated halogenase dataset compared with six commonly used tools (BLASTp, ProteInfer, DeepEC, DEEPre, ECPred, and COFACTOR). This dataset had good diversity covering 11 different EC numbers.

classification framework, contrastive learning could better handle the imbalanced nature of EC numbers, where some EC numbers have thousands of enzyme examples and some only have very few (less than five). In this validation dataset, each type of EC number had no more than five occurrences, and more than 3000 samples were included in this dataset covering more than 1000 different EC numbers. Note that ProteInfer and DeepEC were evaluated using their released pretrained models; thus, our curated validation set appeared during both models' training process. In other words, both ProteInfer and DeepEC had an advantage that both models have seen the validation dataset used in Fig. 2C during training, resulting in the acceptable 0.625 to 0.782 F1 score. Despite this added advantage, CLEAN outperformed both methods, achieving a 0.817 F1 score (Fig. 2C).

We analyzed CLEAN's performance based on the number of times that the EC number occurred in the training set. Even at 50% sequence identity clustering, where the test set and train set had a low similarity, CLEAN's performance did not drop considerably when the number of training examples was scarce (Fig. 2D). With the given results, the two independent datasets (New-392 and Price-149) were combined and revisited. As shown in Fig. 2E, the accuracy performance was studied separately based on the number of times that EC numbers appeared in the training set. As expected, ProteInfer and DeepEC showed a bias toward popular EC numbers, limited by the classification framework. By contrast, CLEAN showed the most superiority in predicting understudied functions and maintained high accuracy regardless of the EC occurrences. The challenge posed by the biased dataset to the classification model was the lack of positive examples for understudied EC numbers. As a result, classification models can hardly learn from the limited positive examples. To further analyze the hypothesis that CLEAN can leverage not only positive examples but also negative examples through contrastive learning, Supcon-Hard loss (SupconH)—a loss function that samples more negatives compared with triplet loss—was implemented (materials and methods; supplementary text, section 2; and fig. S2).

Moreover, we implemented a method to quantify the prediction result confidence. We fitted a two-component Gaussian mixture model (GMM) on the distribution of the Euclidean distances between enzyme sequence embeddings and EC number embeddings (materials and methods). Knowing the prediction confidence, researchers can make quantitative interpretations of CLEAN's prediction. The confidence quantification can also help CLEAN to avoid overprediction by reporting the third level of EC number when the confidence is low (figs. S11 to S14 and supplementary text, section 3).

Experimental validation

Next, we sought to validate the prediction accuracy of CLEAN in assigning EC numbers using halogenases as a proof-of-concept study. Halogenases have been increasingly used for biocatalytic C-H functionalization because of their excellent catalyst-controlled selectivity (*23*, *24*, *25*). Generally, small molecules with halogen atoms produced by halogenases have promising bioactivity and physicochemical properties, thereby offering broad application in pharmaceutical and agrochemical fields (*24*, *26*, *27*). To date, 36 incompletely annotated halogenases have been identified from UniProt, covering all four types of halogenases [haloperoxidase, flavin-dependent, α-ketoglutarate (α-KG)–dependent, and *S*-adenosyl-L-methionine (SAM)–dependent

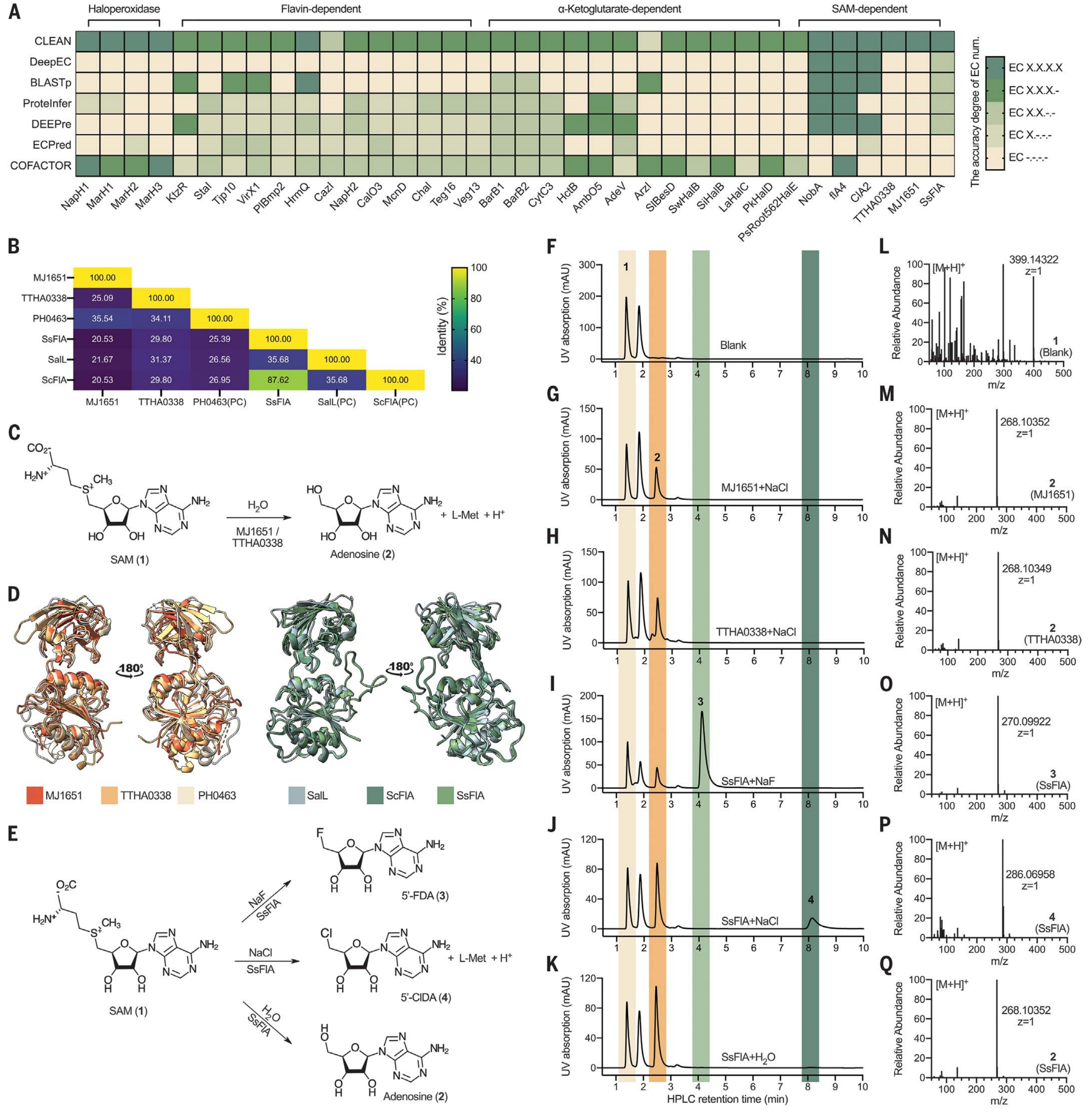


Fig. 3. Experimental validation of CLEAN on uncharacterized halogenases. (**A**) The accuracy degree heatmap of EC numerical ID was shown for the 36 identified halogenases. (**B**) Heatmap of sequence identity among the uncharacterized proteins and positive control (PC) enzymes. The color bar with the "viridis" color scale indicates percentage. (**C**) The SAM hydroxide adenosyltransferase MJ1651-TTHA0338 reaction. (**D**) Structural superposition of the three-dimensional (3D) structures of uncharacterized proteins MJ1651 [Protein Data Bank (PDB) ID: 2F4N (*28*)], TTHA0338 [PDB ID: 2CW5 (*39*)], and positive control enzyme PH0463 [PDB ID: 1WU8 (*40*)]. The same structural superposition was performed for SsFlA [PDB ID: 5B6I (*30*)], SalL [PDB ID: 2Q6O (*41*)], and ScFlA [PDB ID: 1RQR (*42*)]. The superposition shows that the 3D structures of these SAM-binding enzymes are very similar; yet, CLEAN can accurately distinguish their functions. Chain A in each crystal structure was used for structural superposition. (**E**) Nucleophilic substitution of SAM with halide ions or H_2O toward SsFlA. (**F** to **K**) HPLC analysis of reaction mixtures of SAM and NaCl/NaF/H_2O with blank (F), purified MJ1651 (G), purified TTHA0338 (H), and purified SsFlA [(I) to (K)]. The peaks of substrate SAM (**1**), product adenosine (**2**), 5′-fluoro-5′-deoxyadenosine (5′-FDA) (**3**), and 5′-chloro-5′-deoxyadenosine (5′-ClDA) (**4**) were labeled with light yellow, orange, green, and dark green, respectively, which were also aligned at the same retention time. UV, ultraviolet; mAU, milli–absorbance unit. (**L** to **Q**) Mass spectra of compounds obtained from the reaction mixtures: substrate **1** in the blank reaction system (L), adenosine (**2**) in MJ1651 catalyzed reaction (M), adenosine (**2**) in TTHA0338 catalyzed reaction (N), 5′-FDA (**3**) (O), 5′-ClDA (**4**) (P), and adenosine (**2**) (Q). m/z, mass/charge ratio.

halogenase] (Fig. 3A and table S2). These halogenases were either labeled with uncharacterized and/or hypothetical proteins in UniProt or had conflicting annotations in the literature. The halogenase dataset is particularly challenging because the halogenase family is understudied, and only a limited number of halogenases are available in the database. With expert curation and experimental validations showing later, all 36 halogenases were confidentially annotated with EC numbers. Overall, CLEAN achieved much better prediction accuracy (86.7 to 100%; Fig. 2F and Fig. 3A) compared with the six other commonly used computational tools (e.g., ~11.1% in DeepEC and 11.1 to 61.1% in ProteInfer). The latter range corresponds to the prediction accuracy at different digits of EC number (from digit 1 to digit 4). These results demonstrate that CLEAN can distinguish enzyme functions even within the regime of similar biocatalytic reactions.

Among these 36 halogenases, three enzymes named MJ1651, TTHA0338, and SsFlA showed conflicting functions according to the comparison between literature (*28*–*30*) and the description in UniProt. CLEAN predicted new EC numbers in these three cases, suggesting that other potential functions might occur. Therefore, we performed in vitro experiments to validate these predictions. High-performance liquid chromatography–mass spectrometry (HPLC-MS) analysis coupled with enzyme kinetic analysis confirmed that MJ1651 is SAM hydrolase (EC 3.13.1.8), as CLEAN predicted, rather than chlorinase (EC 2.5.1.94) or fluorinase (EC 2.5.1.63), as mislabeled in UniProt and by the selected computational tools used in this work (Fig. 3, C, D, F, G, and M; fig. S3; fig. S4, A and B; fig. S5A; fig. S7; and table S3). CLEAN also correctly annotated TTHA0338, which belongs to the DUF62 Pfam family with no known function, as a SAM hydrolase (Fig. 3, C, D, H, and N; figs. S5B and S7; and table S3). With the exception of BLASTp successfully predicting the target TTHA0338, all other six commonly used computational tools failed to predict MJ1641 and TTHA0338. These results revealed that CLEAN is favorable for correcting mislabeled enzymes and accurately identifying understudied catalytic functions. CLEAN also confidently identified the promiscuous enzyme SsFlA with three EC numbers (EC 2.5.1.63, EC 2.5.1.94, and EC 3.13.1.8; Fig. 3, E, I to K, and O to Q). These observations confirmed that CLEAN could effectively recall defined biological activity and capture elements of enzyme promiscuity. The precision of CLEAN is impressive in distinguishing SAM-binding proteins with homologous structures (fig. S3C) and sequence identity ranging from 20.5 to 35.7% for everything but SsFlA versus ScFlA, which is 87.6% (Fig. 3B and fig. S6). Functions of proteins with sequence identities in this range are often challenging to predict. These results suggest that our sequence-based model CLEAN performed better than structure-based methods [e.g., COFACTOR (*12*, *13*)] in dealing with enzymes with similar structures but different functions.

Discussion

Through systematic in silico and in vitro experimental validations, we have demonstrated that CLEAN achieves superior prediction performance relative to six state-of-the-art tools (i.e., ProteInfer, BLASTp, DeepEC, DEEPre, COFACTOR, and ECPred). A comprehensive analysis on an uncharacterized halogenase dataset indicated that CLEAN can characterize the hypothetical proteins and correct mislabeled proteins, where most sequence-, structure-, and ML-based annotation tools predict incorrectly or are unable to produce a prediction. Identifying enzyme promiscuity is essential for improving the performance of existing enzymes (*3*, *31*), which can be effectively achieved by CLEAN (e.g., SsFlA with three functions). Unlike classification models, contrastive learning is more suitable for biological data, which is usually imbalanced or biased and scarce.

We believe that CLEAN will be a powerful tool for predicting the catalytic function of query enzymes, which can greatly facilitate studies in functional genomics (*32*), enzymology, enzyme engineering (*33*), synthetic biology (*34*), metabolic engineering (*35*, *36*), and retrobiosynthesis (*37*, *38*). Moreover, the general language model representation topped with the contrastive learning workflow used by CLEAN can readily be adapted to other prediction tasks not limited to enzymatic activities, such as functional catalogue (FunCat) and gene ontology (GO). The user-friendly feature of our framework allows CLEAN to be used as an independent tool in a high-throughput manner and as a software component integrated into other computational platforms. The superior performance of CLEAN in predicting understudied proteins should greatly expand the bioinformatics toolbox, thereby laying the cornerstone for future detailed mechanistic studies.

REFERENCES AND NOTES

1. UniProt Consortium, *Nucleic Acids Res.* **49**, D480–D489 (2021).
2. P. Radivojac *et al.*, *Nat. Methods* **10**, 221–227 (2013).
3. K. Hult, P. Berglund, *Trends Biotechnol.* **25**, 231–238 (2007).
4. C. J. Jeffery, *Phil. Trans. R. Soc. B* **373**, 20160523 (2018).
5. E. W. Sayers *et al.*, *Nucleic Acids Res.* **50**, D20–D26 (2022).
6. M. Blum *et al.*, *Nucleic Acids Res.* **49**, D344–D354 (2021).
7. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **215**, 403–410 (1990).
8. D. K. Desai, S. Nandi, P. K. Srivastava, A. M. Lynn, *Adv. Bioinformatics* **2011**, 743782 (2011).
9. S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389–3402 (1997).
10. A. Krogh, M. Brown, I. S. Mian, K. Sjölander, D. Haussler, *J. Mol. Biol.* **235**, 1501–1531 (1994).
11. M. Steinegger *et al.*, *BMC Bioinformatics* **20**, 473 (2019).
12. C. Zhang, P. L. Freddolino, Y. Zhang, *Nucleic Acids Res.* **45**, W291–W299 (2017).
13. A. Roy, J. Yang, Y. Zhang, *Nucleic Acids Res.* **40**, W471–W477 (2012).
14. J. Y. Ryu, H. U. Kim, S. Y. Lee, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 13996–14001 (2019).
15. T. Sanderson, M. L. Bileschi, D. Belanger, L. J. Colwell, *eLife* **12**, e80942 (2023).
16. P. Khosla *et al.*, in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, H. Lin, Eds. (Curran Associates, Inc., 2020), pp. 18661–18673.
17. M. Heinzinger *et al.*, *NAR Genom. Bioinform.* **4**, lqac043 (2022).
18. F. Schroff, D. Kalenichenko, J. Philbin, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2015), pp. 815–823.
19. A. Rives *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016239118 (2021).
20. Y. Li *et al.*, *Bioinformatics* **34**, 760–769 (2018).
21. C. Yu, N. Zavaljevski, V. Desai, J. Reifman, *Proteins* **74**, 449–460 (2009).
22. A. Dalkiran *et al.*, *BMC Bioinformatics* **19**, 334 (2018).
23. M. N. Price *et al.*, *Nature* **557**, 503–509 (2018).
24. C. Crowe *et al.*, *Chem. Soc. Rev.* **50**, 9443–9481 (2021).
25. K. Prakinee *et al.*, *Nat. Catal.* **5**, 534–544 (2022).
26. J. Latham, E. Brandenburger, S. A. Shepherd, B. R. K. Menon, J. Micklefield, *Chem. Rev.* **118**, 232–269 (2018).
27. V. Agarwal *et al.*, *Chem. Rev.* **117**, 5619–5674 (2017).
28. K. N. Rao, S. K. Burley, S. Swaminathan, *Proteins* **70**, 572–577 (2008).
29. A. S. Eustáquio, J. Härle, J. P. Noel, B. S. Moore, *ChemBioChem* **9**, 2215–2219 (2008).
30. H. Sun *et al.*, *Angew. Chem. Int. Ed.* **55**, 14277–14280 (2016).
31. H. Nam *et al.*, *Science* **337**, 1101–1104 (2012).
32. O. Shalem, N. E. Sanjana, F. Zhang, *Nat. Rev. Genet.* **16**, 299–311 (2015).
33. Y. Wang *et al.*, *Chem. Rev.* **121**, 12384–12444 (2021).
34. H. Zhao, *ACS Synth. Biol.* **11**, 3550 (2022).
35. X.-M. Sun, Y.-S. Xu, H. Huang, *Trends Biotechnol.* **39**, 648–650 (2021).
36. G. B. Kim, W. J. Kim, H. U. Kim, S. Y. Lee, *Curr. Opin. Biotechnol.* **64**, 1–9 (2020).
37. T. Yu *et al.*, *Nat. Catal.* **6**, 137–151 (2023).
38. D. Rother, S. Malzacher, *Nat. Catal.* **4**, 92–93 (2021).
39. S. Satoh *et al.*, RIKEN Structural Genomics/Proteomics Initiative (RSGI), Crystal Structure of the Conserved Hypothetical Protein TTHA1091 from *Thermus thermophilus* HB8 (PDB, Entry 1VGG, 2004); http://doi.org/10.2210/pdb1vgg/pdb.
40. K. Shimizu, N. Kunishima, RIKEN Structural Genomics/Proteomics Initiative (RSGI), Crystal structure of project PH0463 from *Pyrococcus horikoshii* OT3 (PDB, Entry 1WU8, 2005); http://doi.org/10.2210/pdb1wu8/pdb.
41. A. S. Eustáquio, F. Pojer, J. P. Noel, B. S. Moore, *Nat. Chem. Biol.* **4**, 69–74 (2008).
42. C. Dong *et al.*, *Nature* **427**, 561–565 (2004).
43. T. Yu *et al.*, Enzyme function prediction using contrastive learning, version 1.0.0, Zenodo (2023); https://doi.org/10.5281/zenodo.7582241.

ACKNOWLEDGMENTS

We thank H. Ren and C. Huang for their suggestions on the characterization of the products formed by halogenases. We also thank for Z. Zhang, C. Huang, and Z. Pei for valuable discussion and guidance on experimental validation. **Funding:** This work was supported by the Molecule Maker Lab Institute: An AI Research Institutes program supported by the US National Science Foundation (NSF) under grant no. 2019897 (H.Z.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the NSF. **Author contributions:** T.Y., Y.L., and H.Z. conceived of the presented idea. T.Y. implemented the computational framework. T.Y., Y.L., H.Z., and J.C.L. designed the in silico experiments. T.Y. and J.C.L. contributed to data preparation and carried out the in silico experiments. J.C.L. contributed to cleaning up code. H.C. and H.Z. planned the in vitro experiments. H.C. and G.J. carried out in vitro experiments and data analysis. T.Y. and H.C. wrote the manuscript with input from all authors. All authors discussed the results and contributed to the final manuscript. H.Z. provided supervision and resources for this study. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data and code generated as part of this study are freely accessible in either the supplementary materials or open repositories. Code

for model development and validation are freely accessible at Zenodo (*43*) and GitHub (https://github.com/tttianhao/CLEAN). CLEAN is converted into an easy-to-use web server and made freely accessible at https://moleculemaker.org/alphasynthesis. The following datasets curated in previous publications and databases were used: Price-149 (*15*) and New-392 (*43*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf2465
Materials and Methods
Supplementary Text
Figs. S1 to S15
Tables S1 to S3
References (*44*–*80*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 8 October 2022; accepted 6 March 2023
10.1126/science.adf2465

Check for updates

ORGANIC CHEMISTRY

Oxidative rearrangement of 1,1-disubstituted alkenes to ketones

Qiang Feng, Qian Wang, Jieping Zhu*

The Wacker process, which is widely used to convert monosubstituted alkenes into the corresponding methyl ketones, is thought to proceed through a Pd^{II}/Pd^{0} catalytic cycle involving a β-hydride elimination step. This mechanistic scenario is inapplicable to the synthesis of ketones from the 1,1-disubstituted alkenes. Current approaches based on *semi*-pinacol rearrangement of Pd^{II} intermediates are limited to the ring expansion of highly strained methylene cyclobutane derivatives. Herein, we report a solution to this synthetic challenge by designing a Pd^{II}/Pd^{IV} catalytic cycle incorporating a 1,2-alkyl/Pd^{IV} dyotropic rearrangement as a key step. This reaction, compatible with a broad range of functional groups, is applicable to both linear olefins and methylene cycloalkanes, including macrocycles. Regioselectivity favors the migration of the more substituted carbon, and a strong directing effect of the β-carboxyl group was also observed.

The Wacker process to convert ethylene to acetaldehyde is one of the most important industrial processes based on transition metal catalysis (*1*, *2*). This aerobic Pd^{II}-catalyzed oxidative process has subsequently been developed into a powerful and reliable method for the conversion of monosubstituted terminal olefins **1** to methyl ketones **2** and is widely used in the synthesis of bioactive molecules (Fig. 1A) (*3*). Mechanistically, coordination of Pd^{II} to alkene followed by Markovnikov hydroxypalladation, β-hydride elimination, and tautomerization accounts for the reaction outcome. Substantial work has been dedicated to reversing the regioselectivity of the hydroxypalladation, and conditions to transform terminal alkenes **1** to aldehydes **3** involving an *anti*-Markovnikov hydroxypalladation step have been reported (*4*). Internal alkenes have also been examined. However, the reaction affords in general a mixture of two regioisomeric ketones unless a neighboring chelating group can direct the hydroxypalladation process (*5*, *6*). Conversely, reports on the Wacker reaction of 1,1-disubstituted olefins remain scarce. Grigg and co-workers reported that methylenecyclobutane was converted to cyclopentanone under typical Wacker conditions ($PdCl_2$-$CuCl_2$-O_2-H_2O) (Fig. 1B) (*7*). Wahl and co-workers have very recently reported optimized conditions using *tert*-butyl nitrite as a terminal oxidant [$PdCl_2(MeCN)_2$-*t*BuONO-EtOH-H_2O, 30°C] (*8*). However, only ring expansion of methylenecyclobutanes to the cyclopentanones was described. Because no β-hydride is available for Pd in complex **4**, a *semi*-pinacol type rearrangement is proposed to be a possible pathway leading to the formation of ring expansion product. Highly toxic reagents such as lead tetraacetate (*9*) and thallium nitrate (*10*, *11*) have been used as stoichiometric oxidants for the same transformation, but again with a very limited substrate scope. Cyanogen azide, a primary explosive, has also been applied for the ring expansion of methylenecycloalkanes (*12*), whereas hydroxy (tosyloxy)iodobenzene (Koser reagent) has been shown to be applicable only to α-substituted styrene derivatives (*13*).

In light of the prevalence of the carbonyl group in bioactive natural products (*14*) and its synthetic versatility (*15*), the development of a general method allowing conversion of 1,1-disubstituted olefins to the rearranged ketones would be broadly useful. To address this challenge, we moved from the Pd^{II}/Pd^{0} catalytic cycle of Wacker conditions to Pd^{II}/Pd^{IV} chemistry (*16*). Our working hypothesis exploiting the dyotropic rearrangement (*17*, *18*) is depicted in Fig. 1C. Hydroxypalladation of alkenes **5** following Markovnikov's rule would afford a reasonably stable neopentyl σ-Pd^{II}

Laboratory of Synthesis and Natural Products (LSPN), Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne, EPFL-SB-ISIC-LSPN, BCH5304, CH-1015 Lausanne, Switzerland.
*Corresponding author. Email: jieping.zhu@epfl.ch

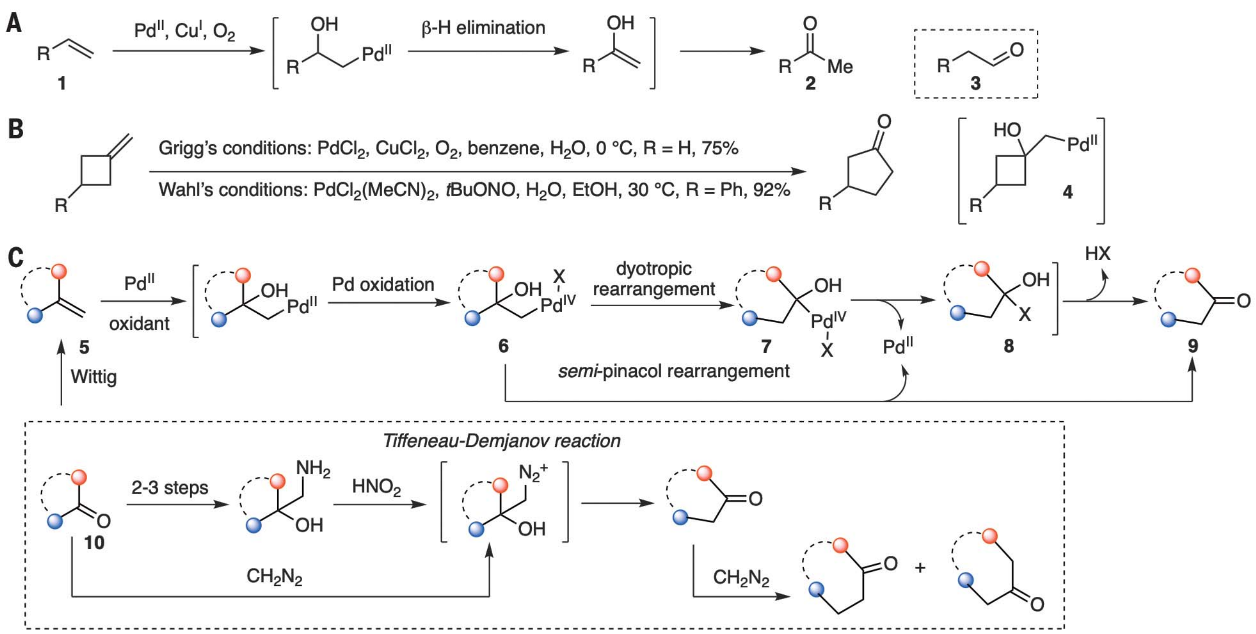


Fig. 1. Oxidative functionalization of alkenes. (**A**) Wacker reaction of monosubstituted alkenes. The β-hydride elimination terminates the Pd^{II}/Pd^{0} catalytic cycle. (**B**) Wacker reaction of 1,1-disubstituted terminal alkenes using Grigg's (*7*) and Wahl's (*8*) previous work on ring expansion of the methylenecyclobutanes. (**C**) Reaction design. A Pd^{II}/Pd^{IV} catalytic cycle involving dyotropic or *semi*-pinacol rearrangement. For comparison purposes, the Tiffeneau-Demjanov homologation of ketones involving diazonium intermediates is presented in the same figure.

complex that would be oxidized in situ to Pd^{IV} species **6** in the presence of an appropriate oxidant. A 1,2-alkyl/Pd^{IV} dyotropic rearrangement of the latter would provide **7** (*19–21*), which, upon C–X bond-forming reductive elimination, would be converted to **8**. Further elimination of HX from **8** would then provide ketone **9**, assuming that X is a reasonably good nucleofuge. Alternatively, a *semi*-pinacol rearrangement (*22*) of Pd^{IV} intermediate **6** could afford the rearranged ketone **9** directly. To streamline the proposed reaction sequence, the choice of oxidant would be of utmost importance; it must satisfy at least the following two criteria: (i) chemoselectively oxidize the in situ generated AlkylPd^{II} species rather than the Pd^{II} salt and (ii) generate a relatively stable alkylPd^{IV}-X species **6** to avoid the premature C–X bond-forming reductive elimination from **6** or a S_N2 substitution reaction (*23*). We report herein the realization of such a transformation that converts diversely functionalized terminal alkenes to the rearranged ketones and methylenecycloalkanes to the ring-expanded cycloalkanones, including macrocycloalkanones. Because terminal alkenes **5** are easily prepared from ketones **10**, the realization of the planned transformation represents a two-step, one-carbon homologation or ring expansion of ketones, a transformation of high synthetic importance. Previously, the Tiffeneau-Demjanov reaction has most often been used for this purpose (*24*). However, three to four steps are needed to accomplish this sequence with moderate overall synthetic efficiency. Alternatively, reaction of ketones with diazomethane has also been used for one-step homologation (*25*). This latter reaction is often complicated by the competitive formation of epoxide and multiple methylene insertion process. In addition, both reactions involve the use of unstable and hazardous reagents (HNO_2 and CH_2N_2) and are applied mainly to the ring expansion of cyclic ketones.

Reaction development

Using 7-methylenetridecane (**11a**, R = nC_6H_{13}) as a test substrate, conditions were surveyed by varying systematically the Pd(II) sources, the oxidants, the temperature, the solvents, and the ligands (tables S1 to S6). Performing the reaction in $MeCN/H_2O$ (v/v = 4:1) at room temperature in the presence of Selectfluor {1-chloromethyl-4-fluoro-1,4-diazoniabicyclo[2.2.2] octane bis(tetrafluoroborate); 1.2 equiv} and a catalytic amount of $Pd(MeCN)_4(BF_4)_2$ (10 mol%) afforded the rearranged ketone **12a** in 75% yield. Other fluorine-based oxidants, such as diversely substituted *N*-fluoropyridinium salts and *N*-fluorobenzenesulfonamide, were inefficient, affording **12a** in less than 10% yield at best. The counterion BF_4^- in Selectfluor was also important as changing it to PF_6^- afforded

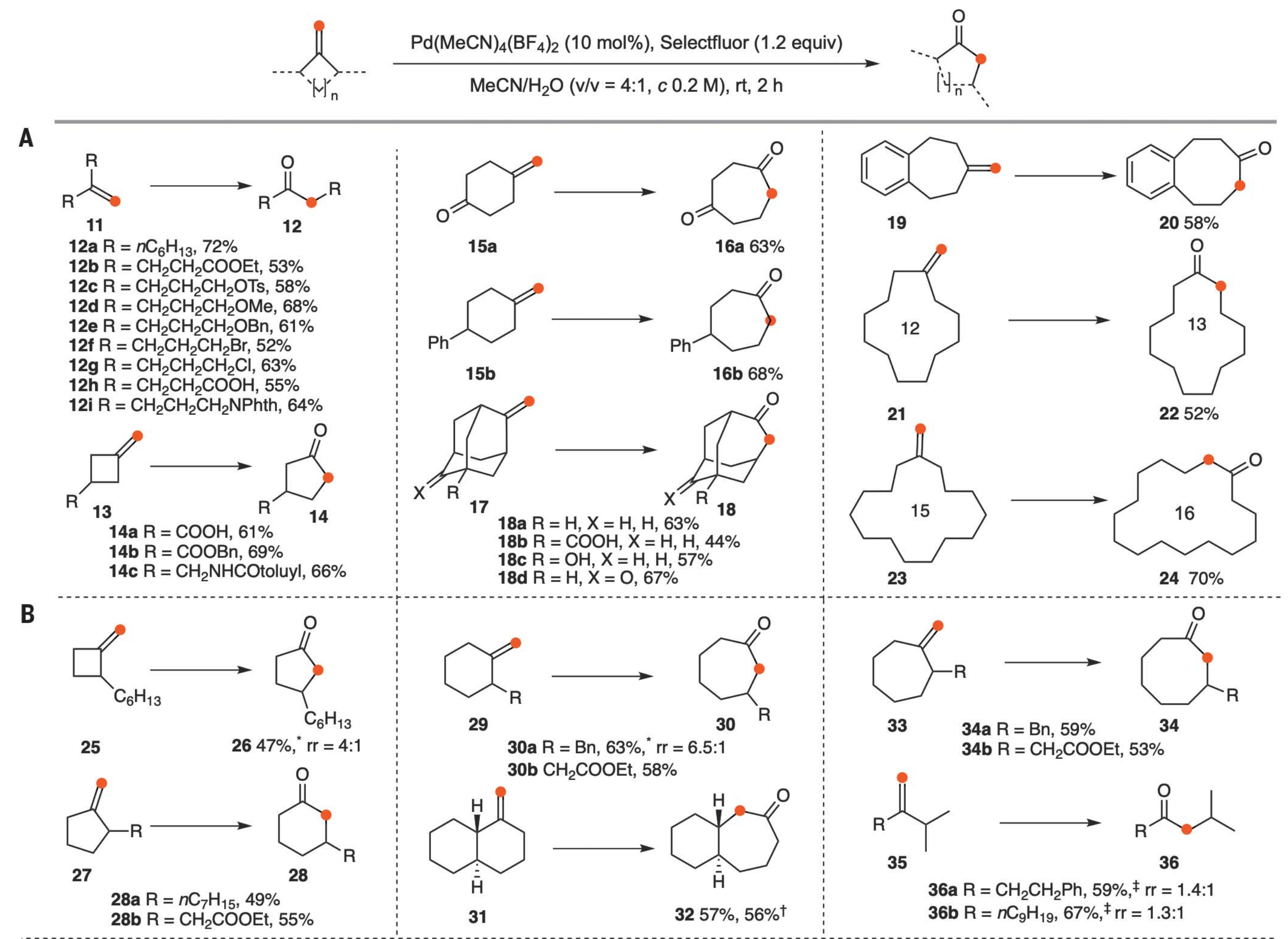


Fig. 2. From 1,1-disubstituted alkenes to rearranged ketones. Structure of the minor isomer is not shown. (**A**) Symmetric 1,1-disubstituted alkenes. (**B**) Unsymmetrically substituted methylene cycloalkanes. Reactions were performed at 0.2 mmol scale, and yields refer to the isolated products unless otherwise specified. *Yield of pure regioisomer. †Reaction performed at 5 mmol scale. ‡Yield of two inseparable regioisomeric ketones. The rr was determined by ^1H NMR spectral integration of the crude product using mesitylene as an internal standard.

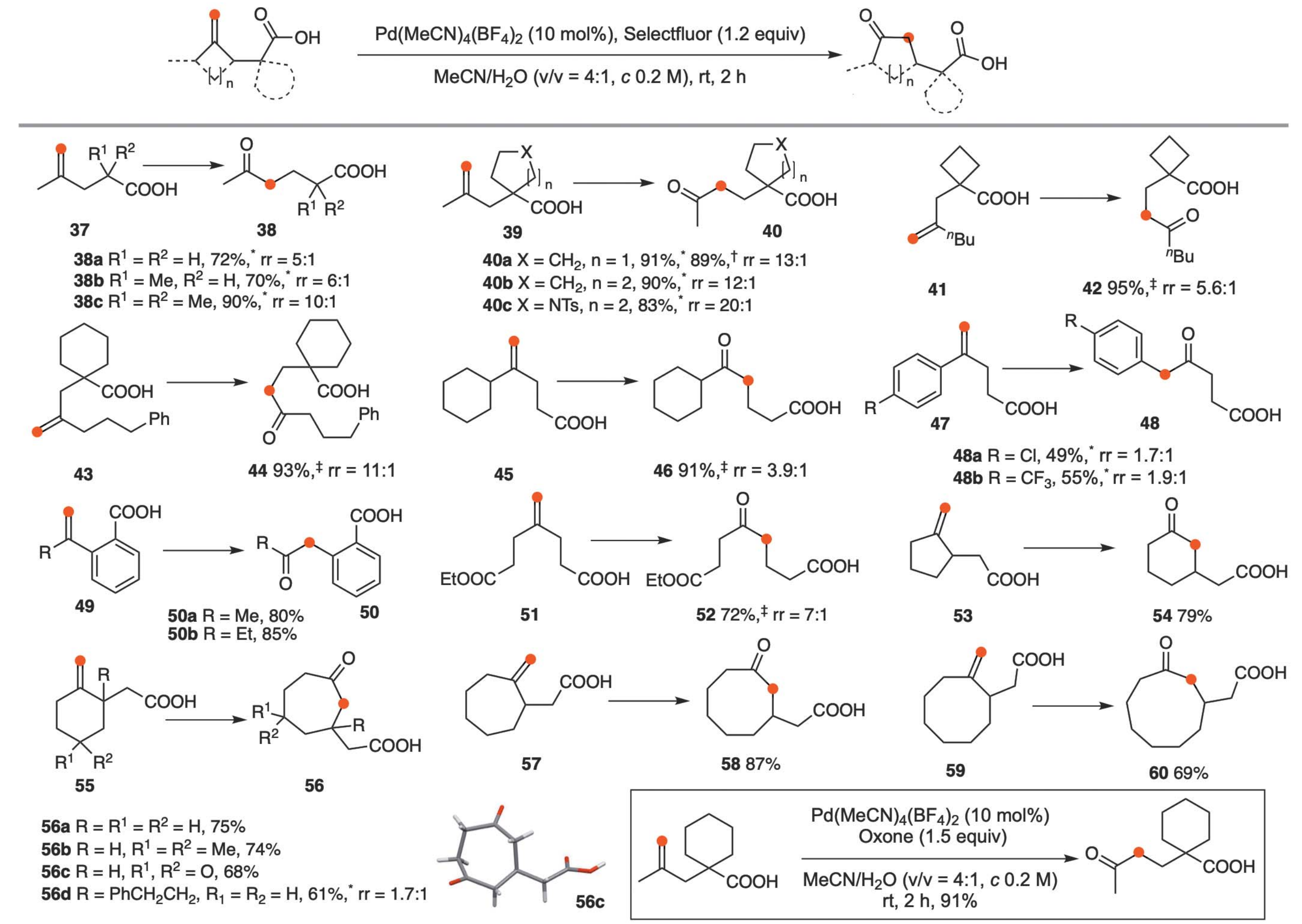


Fig. 3. Directing effect of carboxyl group. *Yield of pure regioisomer. †Reaction performed at 6 mmol scale in the presence of 1.0 mol % of Pd catalyst. ‡Yield of two inseparable regioisomeric ketones.

12a in a diminished yield (58%). Bis(trifluoroacetoxy)iodo benzene can also promote this oxidative transformation, furnishing **12a** in 29% yield (table S1). Isolation of 2-hexyloctane-1,2-diol (16%) and 2-hexyl-2-hydroxy-octyl trifluoroacetate (12%) in this case indicated clearly the important role of the counterion (X) in determining the reaction pathway of Pd^{IV} species (compare intermediate **6**; Fig. 1C). Other Pd^{II} salts, such as $Pd(OAc)_2$, Pd $(OCOCF_3)_2$, and $PdCl_2$, were less efficient catalysts than $Pd(MeCN)_4(BF_4)_2$ (table S2). The higher Lewis acidity of the latter might be beneficial to the metal's coordination to the double bond and the subsequent hydroxypalladation reaction. Bipyridine-based ligands were detrimental to the reaction (table S6).

Reaction scope

With the optimum conditions thus determined, the reaction scope was examined (Fig. 2A). A range of symmetric 1,1-disubstituted alkenes bearing diverse functional groups such as ester, tolylsulfonate, alkyl ethers (MeO, BnO), alkyl halides (Cl, Br), carboxylic acid, and phthalimide and hydroxyl groups, were successfully transformed to the rearranged ketones (**12b** to **12i**) in good yields. Beyond one-carbon ring expansion of strained methylenecyclobutanes (**13a** to **13c**) affording the cyclopentanones **14a** to **14c** as expected, ring expansion of functionalized methylenecyclohexanes (**15a** and **15b**) proceeded equally well to afford seven-membered cycloheptanones **16a** and **16b** in good yields. Functionalized methyleneadamantanones **17a** to **17d** (R = COOH and OH, X = H, H and O) were likewise transformed to homoadamantanones **18a** to **18d**. Benzocyclooctan-7-one (**20**), which is difficult to access by conventional cyclization methods, was obtained in good yield through ring expansion of 7-methylenebenzocycloheptane (**19**). Finally, ring expansion of macrocycles proceeded smoothly, generating the 13- and 16-membered cycloalkanones **22** and **24** from the 12- and 15-membered methylenecycloalkanes **21** and **23**, respectively, in good yields.

Unsymmetric 1,1-disubstituted alkenes were next examined (Fig. 2B). The 2-substituted cycloalkanon-1-ones were in general converted to a mixture of two ring-enlarged ketones in low selectivity by the Tiffeneau-Damjanov reaction and by the diazomethane addition method (*26*). We were therefore delighted to observe that the 1-alkyl-2-methylenecycloalkanes underwent ring expansion to the corresponding C3-substituted cyclopentanones (**26**), cyclohexanones (**28**), and cycloheptanones (**30**) resulting from selective migration of the more substituted carbon. *Trans*-1-methylenedecalin (**31**) was converted regioselectively to *trans*-decahydro-6*H*-benzo[7]annulen-6-one (**32**) in 57% yield. The reaction of *trans*-decalin-1-one with diazomethane afforded a complex reaction mixture of regioisomeric ketones and epoxide (*27*). Finally, the 3-substituted cyclooctanones **34a**

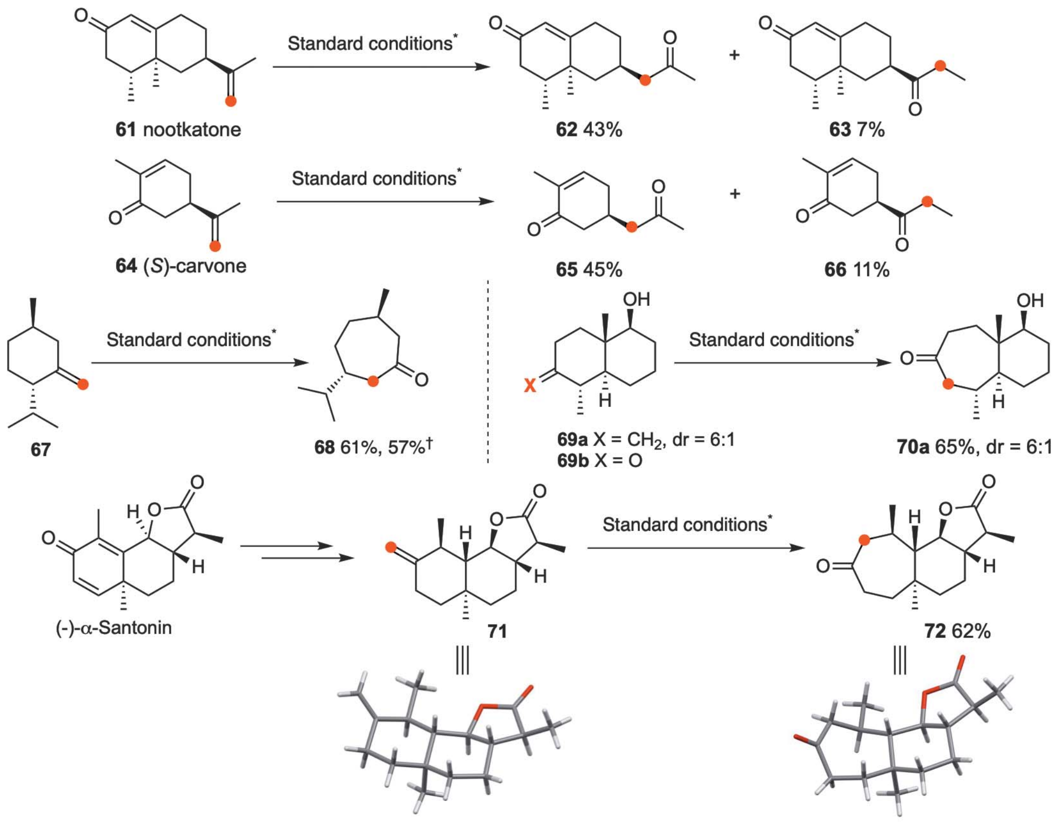


Fig. 4. Application to natural products and their derivatives. *Standard conditions: $Pd(MeCN)_4(BF_4)_2$ (10 mol %), Selectfluor (1.2 equiv), and $MeCN/H_2O$ (v/v = 4:1, *c* 0.2 M) at room temperature for 2 hours. †Reaction performed at 6 mmol scale. dr, diastereomeric ratio.

and **34b** were obtained exclusively from the 1-substituted 2-methylenecycloheptanes **33a** and **33b**. However, linear unsymmetric terminal alkenes **35** were oxidatively rearranged to ketones **36** in low selectivity.

The low regioselectivity observed with the linear unsymmetric terminal alkenes **35** prompted us to examine the directing effect of common organic functional groups. Being easily accessible and transformable, carboxylic acid was deemed to be an ideal candidate. When 4-methylpent-4-enoic acid (**37a**) was submitted to the standard conditions, 5-oxohexanoic acid (**38a**, 72%) resulting from the migration of the carboxylethyl group was formed as a major product (*28*). Introducing substituents α to the carboxylic acid increased the regioselectivity (**38b** and **38c**): **38c** could be isolated in 90% yield [regioisomeric ratio (rr) = 10:1]. Similarly, cyclopentanecarboxylic acid, cyclohexanecarboxylic acid, and *N*-Ts piperidine-3-carboxylic acid derivatives (**39a** to **39c**) were rearranged to the corresponding δ-oxocarboxylic acids (**40a** to **40c**) in excellent yields and regioselectivities. When the two substituents of the Csp^2 were both secondary alkyl groups as in **41** and **43**, the one bearing the carboxylic acid functionality again displayed higher migratory aptitude leading to **42** and **44**, respectively, in excellent yields and selectivities. Migration of the secondary CH_2CH_2COOH group dominated over the tertiary alkyl group (cyclohexyl) as seen in the conversion of **45** to **46**. Styrene derivatives **47a** and **47b** were converted to a mixture of two regioisomeric ketones, **48a** and **48b**, with low selectivities caused by the intrinsically high migratory aptitude of the phenyl group. However, styrene derivatives **49a** and **49b** were converted to **50a** and **50b** exclusively resulting from the preferential migration of the 2-carboxylphenyl group. Apparently, the high migratory aptitude of the phenyl group and the directing effect of the carboxylic group acted synergistically in this case. The strong directing effect of the carboxylic acid over an ester was evidenced in the conversion of **51** to **52** (rr = 7:1). Ring expansion of 2-(2-methylenecycloalkanyl)acetic acids proceeded selectively to provide one-carbon homologated 3-substituted cyclohexanone **54**, cycloheptanones **56**, cyclooctanone **58**, and cyclononanone **60** as the only products in high yields. Quaternary carbon displayed a diminished migratory aptitude as compound **56d** was isolated in 61% yield, together with its regioisomer (35%, rr = 1.7:1), albeit in excellent overall yield (96%). Finally, performing the reaction of **39a** at 6 mmol scale in the presence of only 1.0 mol % of $Pd(MeCN)_4(BF_4)_2$ afforded **40a** in 89% yield, indicating the practicality of the present protocol.

Oxone proved to be a competent oxidant for the oxidative rearrangement of **39b**, affording **40b** in 91% yield (Fig. 3, inset, and tables S8 and S9). Nevertheless, applying the same conditions to **11a** afforded product **12a** in only 8% yield. Selectfluor remained therefore a superior oxidant in terms of generality.

Application to natural products and derivatives

Under the standard conditions, nootkatone (**61**) was converted to two isomeric ketones, **62** (43%) and **63** (7%, rr = 6:1), whereas (*S*)-carvone (**64**) was transformed to **65** (45%) and **66** (11%, rr = 4:1) (Fig. 4). Ring expansion of **67**, which was derived from L-menthone, and bicyclic compound **69a** provided **68** (61%) and **70a** (65%), respectively, resulting from the selective migration of the tertiary alkyl groups. Using the diazomethane method, the hydroxyl group in ketone **69b** has to be protected to avoid decomposition, and a mixture of two rearranged ketones was formed under various conditions (*29*). Finally, an α-santonin derivative **71** underwent regioselective ring enlargement to furnish **72** in 62% yield. As seen from the x-ray structures of **71** and **72**, the alkyl group shifted with retention of its absolute configuration, indicating a concerted migration process.

Mechanistic studies

A series of control experiments was performed to gain insight into the reaction mechanism. In the absence of Selectfluor under otherwise standard conditions, compound **39a** was partially

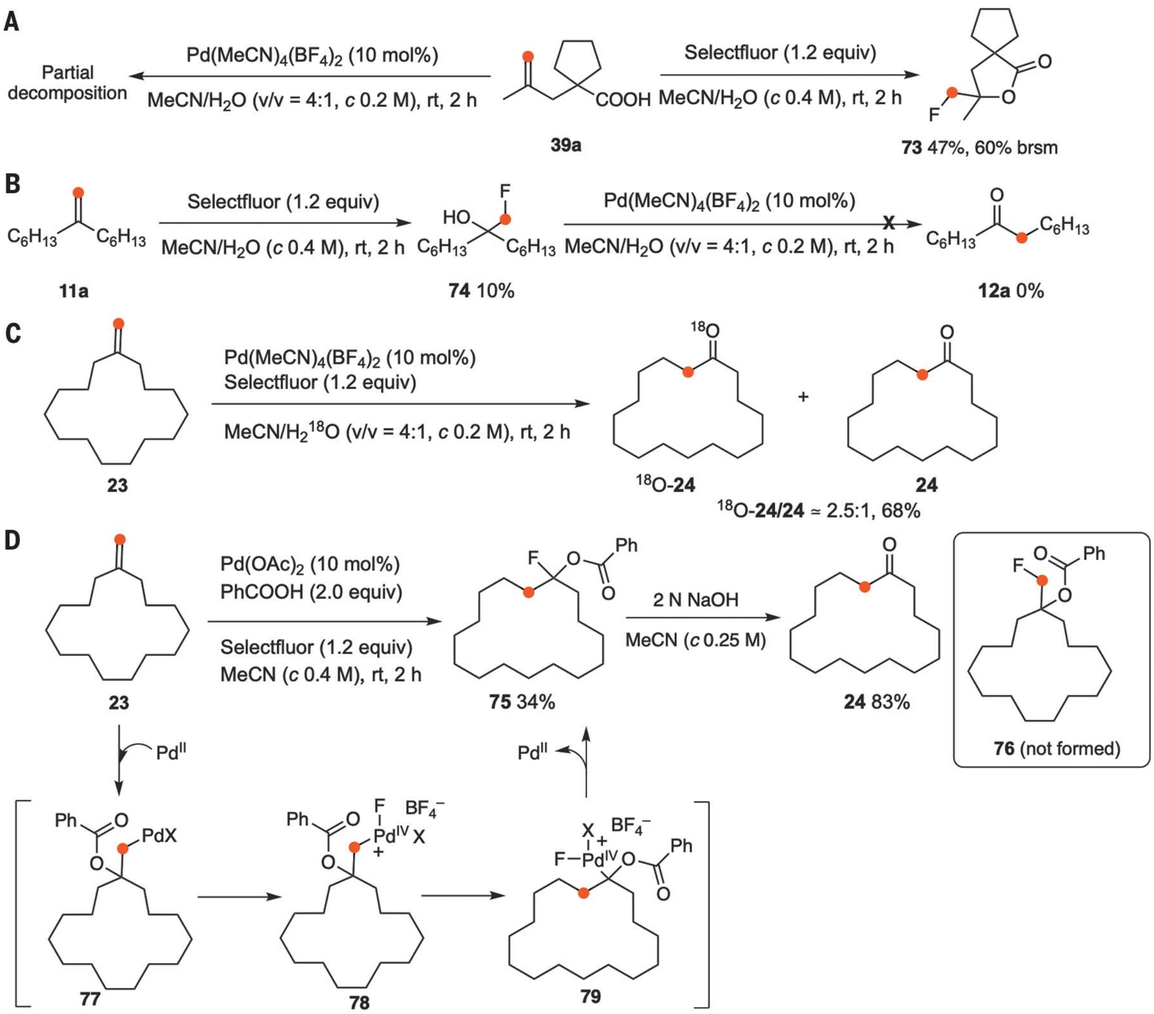


Fig. 5. Results of control experiments. (**A**) Both Selectfluor and $Pd(MeCN)_4(BF_4)_2$ are needed for the oxidative rearrangement. brsm, based on recovered starting material. (**B**) 7-(fluoromethyl)tridecan-7-ol (**74**) is not the intermediate en route to the rearranged ketone **12a**. (**C**) ^{18}O labeling indicates that water is the source of the ketone oxygen. (**D**) Isolation of compound **75** provides evidence that 1,2-alkyl/Pd^{IV} dyotropic rearrangement might be operational.

decomposed without formation of the rearranged ketone **40a**. On the other hand, stirring a $MeCN/H_2O$ solution of **39a** with Selectfluor at room temperature in the absence of Pd salt afforded fluorolactonization product **73** in 47% yield (Fig. 5A) (*30*). Submitting 7-(fluoromethyl) tridecan-7-ol (**74**), prepared through a Selectfluor-mediated intermolecular hydroxyfluorination of alkene **11a** (*31*), failed to afford the rearranged ketone (Fig. 5B). The results of these three experiments indicate that both Selectfluor and Pd(II) salt are needed for the present reaction and that fluorinated alcohol **74** is not an intermediate en route to the rearranged ketone **12a**. No reaction took place between Selectfluor and $Pd(MeCN)_4(BF_4)_2$ at room temperature as evidenced by ^1H and ^{19}F nuclear magnetic resonance (NMR) spectra, suggesting that Pd^{IV} species is formed only after the generation of an alkylPd^{II} intermediate. Replacing H_2O by ^{18}O-labeled H_2O (97% ^{18}O) under otherwise standard conditions, methylenecyclopentadecane (**23**) was converted to cyclohexadecanone (^{18}O-**24**/**24**-ratio 2.5:1) in 68% yield, suggesting that water is the source of oxygen in this transformation (Fig. 5C). Hydration of ketone during the work-up and column chromatographic purification might be responsible for the lower than expected ^{18}O incorporation in **24**.

Finally, stirring an acetonitrile solution of **23**, benzoic acid (2.0 equiv), and Selectfluor (1.2 equiv) in the presence of a catalytic amount of $Pd(OAc)_2$ (0.1 equiv) afforded the ring expansion product **75** (34% yield), which, upon saponification, was converted to **24**. Compound **76** resulting from fluoropalladation of alkene was not observed under these conditions (*32*). The formation of **75** could be accounted for by the following reaction sequence: A Pd^{II}-catalyzed acyloxypalladation of alkene **23** (*33*, *34*) would afford Pd^{II} intermediate **77**, which upon Pd oxidation, would be converted to Pd^{IV} species **78** (*35*), which is stable enough to engage in the dyotropic rearrangement leading to **79**. The C–F bond-forming reductive elimination from the latter would furnish **75** with concurrent regeneration of the Pd^{II} catalyst (Fig. 5D). The results of these control experiments indicate that dyotropic rearrangement could be a possible reaction pathway in the oxidative conversion of the 1,1-disubstitued alkenes to the rearranged ketones, although *semi*-pinacol rearrangement of intermediate **6** cannot be eliminated at this stage (Fig. 1C).

In summary, we have developed a Pd^{II}-catalyzed oxidative conversion of 1,1-disubstituted alkenes to the rearranged ketones. The reaction, applicable to both the linear terminal alkenes and the methylenecycloalkanes, is compatible with a wide range of functional groups including alkyl halides, aryl halides, alkyl tosylate, hydroxyl, carboxylic acid, ester, lactone, amide, and ketone and α,β-unsaturated ketone, providing products with handles for further chemical transformations.

REFERENCES AND NOTES

1. J. Smidt *et al*., *Angew. Chem.* **71**, 176–182 (1959).
2. J. Tsuji, *Synthesis* **1984**, 369–384 (1984).
3. R. A. Fernandes, A. K. Jha, P. Kumar, *Catal. Sci. Technol.* **10**, 7448–7470 (2020).
4. J. J. Dong, W. R. Browne, B. L. Feringa, *Angew. Chem. Int. Ed.* **54**, 734–744 (2015).
5. B. Morandi, Z. K. Wickens, R. H. Grubbs, *Angew. Chem. Int. Ed.* **52**, 2944–2948 (2013).
6. R. J. DeLuca *et al*., *J. Org. Chem.* **78**, 1682–1686 (2013).

7. P. Boontanonda, R. J. Grigg, *J. Chem. Soc. Chem. Commun.* **17**, 583–584 (1977).
8. J. Sietmann, M. Tenberge, J. M. Wahl, *Angew. Chem. Int. Ed.* **62**, e202215381 (2023).
9. Y. Matsubara, Y. Youzhu, C. Zhi, S.-I. Takekuma, H. Yamamoto, *Bull. Chem. Soc. Jpn.* **64**, 2578–2580 (1991).
10. P. Abley, J. E. Byrd, J. Halpern, *J. Am. Chem. Soc.* **95**, 2591–2596 (1973).
11. D. Farcasiu, P. V. R. Schleyer, D. B. Ledlie, *J. Org. Chem.* **38**, 3455–3459 (1973).
12. J. E. McMurry, A. P. Coppolino, *J. Org. Chem.* **38**, 2821–2827 (1973).
13. M. W. Justik, G. F. Koser, *Tetrahedron Lett.* **45**, 6159–6163 (2004).
14. P. Ertl, T. Schuhmann, *J. Nat. Prod.* **82**, 1258–1263 (2019).
15. D. J. Foley, H. Waldmann, *Chem. Soc. Rev.* **51**, 4094–4120 (2022).
16. P. Sehnal, R. J. K. Taylor, I. J. S. Fairlamb, *Chem. Rev.* **110**, 824–889 (2010).
17. M. T. Reetz, *Angew. Chem. Int. Ed.* **11**, 129–130 (1972).
18. I. Fernández, F. P. Cossío, M. A. Sierra, *Chem. Rev.* **109**, 6687–6711 (2009).
19. J. Cao, H. Wu, Q. Wang, J. Zhu, *Nat. Chem.* **13**, 671–676 (2021).
20. G. Yang *et al*., *J. Am. Chem. Soc.* **144**, 14047–14052 (2022).
21. J. Gong, Q. Wang, J. Zhu, *Angew. Chem. Int. Ed.* **61**, e202211470 (2022).
22. Z.-L. Song, C.-A. Fan, Y.-Q. Tu, *Chem. Rev.* **111**, 7523–7556 (2011).
23. G. Liu, S. S. Stahl, *J. Am. Chem. Soc.* **128**, 7179–7181 (2006).
24. S. M. Kohlbacher, V.-S. Ionasz, L. Ielo, V. Pace, *Monatsh. Chem.* **150**, 2011–2019 (2019).
25. N. R. Candeias, R. Paterna, P. M. P. Gois, *Chem. Rev.* **116**, 2937–2981 (2016).
26. R. Hennig, P. Metz, *Angew. Chem. Int. Ed.* **48**, 1157–1159 (2009).
27. C. D. Gutsche, H. H. Peter, *J. Am. Chem. Soc.* **77**, 5971–5977 (1955).
28. B. S. Matsuura *et al*., *Org. Lett.* **13**, 6320–6323 (2011).
29. F. A. Maciás, C. Carrera, N. Chinchilla, F. R. Fronczek, J. C. G. Galindo, *Tetrahedron* **66**, 4125–4132 (2010).
30. D. Parmar, M. S. Maji, M. Rueping, *Chemistry* **20**, 83–86 (2014).
31. G. S. Lal, *J. Org. Chem.* **58**, 2791–2796 (1993).
32. H. Peng, Z. Yuan, H.-Y. Wang, Y.-L. Guo, G. Liu, *Chem. Sci.* **4**, 3172–3178 (2013).
33. R. C. Larock, T. R. Hightower, *J. Org. Chem.* **58**, 5298–5300 (1993).
34. R. M. Trend, Y. K. Ramtohul, B. M. Stoltz, *J. Am. Chem. Soc.* **127**, 17778–17788 (2005).
35. K. M. Engle, T.-S. Mei, X. Wang, J.-Q. Yu, *Angew. Chem. Int. Ed.* **50**, 1478–1491 (2011).

ACKNOWLEDGMENTS

We thank F. Fadaei-Tirani and R. Scopelliti for the x-ray structural analysis of compounds **56c**, **71**, and **72**. **Funding:** This work was supported by the Ecole Polytechnique Fédérale de Lausanne (EPFL) and the Swiss National Science Foundation. **Author contributions:** Q.F., Q.W., and J.Z. conceived the work, designed the experiments, and analyzed the data. Q.F. optimized the reaction conditions and performed the experiments. All authors co-wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Deposits CCDC 2244382 (**56c**), CCDC 2244379 (**71**), and CCDC 2224381 (**72**) contain the crystallographic data for this study and can be obtained free of charge from The Cambridge Crystallographic Data Centre at www.ccdc.cam.ac.uk/data_request/cif. Other characterization data are available in the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg3182
Materials and Methods
Supplementary Text
Tables S1 to S15
NMR Spectra
References (*36*–*77*)

Submitted 15 December 2022; accepted 7 March 2023
10.1126/science.adg3182

REVIEW SUMMARY

MOIRÉ PHOTONICS

Moiré photonics and optoelectronics

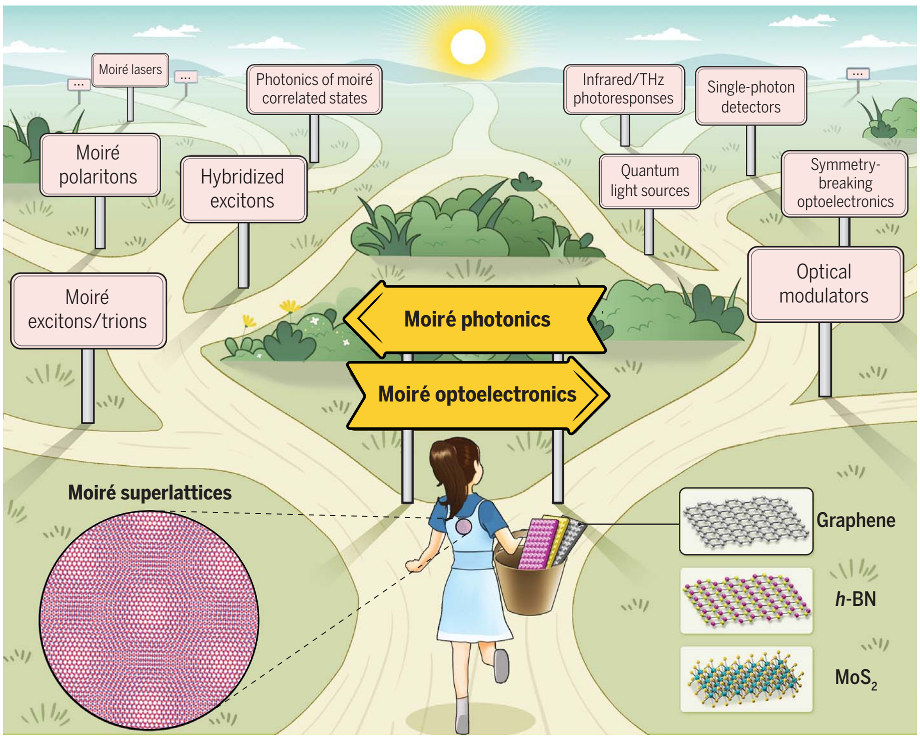
Luojun Du*, Maciej R. Molas, Zhiheng Huang, Guangyu Zhang*, Feng Wang, Zhipei Sun*

BACKGROUND: Moiré pattern refers to a new, long-length scale of structure arising from the interference between two or more periodic templates. The scientific and engineering value of moiré patterns was firstly appreciated by the physicist Lord Rayleigh in 1874. Thereafter, the study of moiré patterns enabled the birth of moiré techniques for a wide variety of practical applications, e.g., precision measurement and positioning, automatization, and strain measurements. In condensed-matter physics, moiré patterns (also referred to as moiré superlattices) can be formed by vertically stacking two or more two-dimensional (2D) layered materials with a tiny twist angle and/or a slight lattice mismatch. Over the past decade, moiré superlattices have made a spectacular appearance with a cornucopia of previously unachieved new phenomena and distinct functionalities, transforming the landscape of fundamental solid-state physics, materials science, and engineering.

ADVANCES: Moiré superlattices introduce a new length-scale potential much larger than the crystal periodicity of the constituent 2D layers, providing a powerful strategy for engineering electronic band structures and hence resulting in a plethora of emergent quantum phenomena. For example, the long-wavelength moiré periodic potential can fold the electronic band structure into a mini-Brillouin zone, giving rise to the formation of flat bands and a rich phase diagram of correlated states, such as unconventional superconductivity, orbital magnetism, Mott insulator states, and topological multiferroic orders.

When moiré superlattices meet light, they provide a powerful platform for discovering emergent photonic and optoelectronic phenomena and device architectures. In early 2019, four research groups independently demonstrated the pronounced effects of moiré potential on excitons of transition-metal dichalcogenide heterostructures from the perspective of both real and momentum spaces. The new degree of freedom afforded by using moiré superlattices provides possibilities for engineering the exciton band structures and light-matter interactions for numerous applications, such as versatile quantum light sources, new quantum many-body physics, and long-sought boson crystals. These breakthroughs in the study of moiré excitons have excited tremendous interest in moiré photonics and optoelectronics. Indeed, a rich diversity of emergent moiré photonic and optoelectronic properties have been witnessed with unprecedented speed over the past few years—properties covering collective moiré excitations, moiré polaritons, photonics of moiré correlated electronic states, giant mid- and far-infrared photoresponses, tunable bulk photovoltaic effect, and so forth. Notably, the moiré bulk photovoltaic effects are ultrastrong and much larger than previous demonstrations, enabling the demonstration of intelligent infrared sensors with merely a subwavelength footprint potentially for next-generation nonlinear photonics and optoelectronics.

OUTLOOK: Currently, all measurements of moiré photonics and optoelectronics are performed in the far-field limit with a collection of signals from more than 10,000 moiré cells. Therefore, developing advanced techniques that can simultaneously probe the photonic and optoelectronic properties in an individual moiré supercell and evaluate the moiré potential landscape will be valuable for advancing our understanding. Meanwhile, current research efforts on moiré photonics and optoelectronics mainly focus on twisted graphene and twisted transition-metal dichalcogenides. Exploring new moiré systems, such as moiré superlattices composed of 2D ferroelectric, magnetic, or multiferroic crystals, and moiré devices involving two or multiple single moiré superlattices, will initiate research directions of moiré photonics and optoelectronics. Furthermore, engineering the moiré photonics and optoelectronics through external means (e.g., electric or magnetic field, strain, and ultrafast optical excitations) has the potential to spark the next "gold rush" for technical innovations (e.g., quantum nonlinear optics, ultracompact optical modulators, and terahertz single-photon devices). We are just starting down the path of exploring moiré photonics and optoelectronics, and there will be many more surprises to come. ■



Moiré photonics and optoelectronics. Moiré superlattices introduce a paradigm to engineer the band structures of collective excitations and light-matter interaction for a plethora of emergent photonic and optoelectronic phenomena, such as moiré excitons, moiré polaritons, photonics of moiré correlated electronic states, intelligent infrared sensing, quantum light sources, single-photon detection, and symmetry-breaking optoelectronics.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: zhipei.sun@aalto.fi (Z.S.); luojun.du@iphy.ac.cn (L.D.); gyzhang@iphy.ac.cn (G.Z.)
Cite this article as L. Du *et al.*, *Science* 379, eadg0014 (2023). DOI: 10.1126/science.adg0014

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adg0014

REVIEW

MOIRÉ PHOTONICS

Moiré photonics and optoelectronics

Luojun Du[1,2,3]*, Maciej R. Molas[4], Zhiheng Huang[2,3], Guangyu Zhang[2,3,5]*, Feng Wang[6], Zhipei Sun[1,7]*

Moiré superlattices, the artificial quantum materials, have provided a wide range of possibilities for the exploration of completely new physics and device architectures. In this Review, we focus on the recent progress on emerging moiré photonics and optoelectronics, including but not limited to moiré excitons, trions, and polaritons; resonantly hybridized excitons; reconstructed collective excitations; strong mid- and far-infrared photoresponses; terahertz single-photon detection; and symmetry-breaking optoelectronics. We also discuss the future opportunities and research directions in this field, such as developing advanced techniques to probe the emergent photonics and optoelectronics in an individual moiré supercell; exploring new ferroelectric, magnetic, and multiferroic moiré systems; and using external degrees of freedom to engineer moiré properties for exciting physics and potential technological innovations.

Two-dimensional (2D) layered materials have ushered in a new era of fundamental research and technological innovation, owing to the electronic, photonic, and optoelectronic properties that are unattainable in their bulk counterparts (*1*–*3*). Today, over 2000 atomically thin 2D materials have been uncovered, ranging from wide-bandgap insulators (e.g., *h*-BN), semiconductors (e.g., MoS_2), and polar metals (e.g., Ga) to superconductors (e.g., $NbSe_2$), ferromagnets (e.g., CrI_3), and quantum spin liquids (e.g., $RuCl_3$) (*1*, *4*–*7*). Notably, 2D atomic layers with diverse properties can be stacked together to form van der Waals (vdW) heterostructures without the constraint of lattice matching in conventional heterostructures. This enables opportunities to combine the best characteristics of different ingredients in one ultimate synthetic quantum material and thus leads to the realization of numerous electronic, photonic, magnetic, and topological functionalities that were previously impossible (*1*, *2*, *8*, *9*).

Intriguingly, a geometrical moiré superlattice emerges as a result of the interference between the constituent 2D atomic layers with a slight lattice mismatch and/or a small rotational twist (Fig. 1A) (*10*, *11*). This moiré superlattice introduces a new length and energy scale, and provides a platform to engineer the band structures (including both single-particle states and collective excitations, Fig. 1B) and the light-matter interaction for exotic quantum phenomena (*12*–*16*). Triggered by the discovery of correlated insulating states and unconventional superconductivity in magic-angle twisted bilayer graphene (TBG) (*17*, *18*), a plethora of exciting electronic, optical, and optoelectronic properties have been discovered in moiré superlattices (Fig. 1C), including moiré excitons (*19*–*22*), versatile quantum light sources (*23*), orbital ferromagnetism (*24*–*26*), Wigner crystal states (*11*, *27*, *28*), stripe phases (*29*, *30*), topological multiferroic order (*31*), boson exciton crystals (*32*), and intelligent infrared sensing (*33*).

We present an overview of the recent progress on emerging moiré photonics and optoelectronics, such as moiré neutral and charged excitons, resonantly hybridized excitons, moiré polaritons, emergent optical responses of moiré correlated electronic states, reconstructed collective excitations, terahertz single-photon detection, strong mid- and far-infrared photoresponses, symmetry-breaking optoelectronics, etc. We also provide an outlook on the major challenges and future opportunities in this field, as well as the implications for potential new technological innovations. With a particular focus on moiré photonics and optoelectronics, our aim is to complement other recent reviews that discuss the fabrication (*34*), correlated electronic states (*10*, *13*), and excitonic physics (*12*, *35*, *36*).

Fabrication and visualization of moiré superlattices

The controlled fabrication and direct visualization of moiré superlattices are essential for unlocking their intriguing electronic and optical properties and future on-demand application development.

Fabrication of moiré superlattices

Top-down methods

One top-down method for preparing high-quality homobilayer moiré superlattices is the "tear-and-stack" technique (*37*, *38*). This technique is based on the deterministic pick-up and transfer, and consists of selectively picking up a section of a 2D material and then transferring it onto the remaining section, which has been rotated by a user-designed angle (*34*, *38*). Because the two constituent layers initially have exactly the same crystal orientation, the twist angle can be controlled with a subdegree resolution. By predetermining the principal crystallographic directions of 2D materials, moiré heterostructures with controllable twist angles can also be fabricated via the deterministic pick-up and transfer technique (*34*). Until now, we have had at our disposal a wide variety of moiré homo- and heterostructures, such as twisted transition-metal dichalcogenide (TMDC) homobilayers (*39*, *40*), twisted bilayer CrI_3 (*41*, *42*), $MoSe_2$/WSe_2 (*19*, *21*), and WS_2/WSe_2 (*20*, *43*). It is noteworthy that an automated robotic pick-up and transfer technique has been recently developed, enabling the rapid manufacture of moiré superlattices with deliberate design and angle control (*44*, *45*). We envision that the advanced robotic assembly, through integration with wafer-scale 2D material growth, machine learning, and computer-vision algorithms (*46*, *47*), could help realize the full potential of moiré superlattices for new physics and technological advances in photonics and optoelectronics.

Bottom-up methods

Developing direct growth methods for synthesizing large-scale and uniform moiré superlattices is important for future technological applications. One of the effective bottom-up methods is the vdW epitaxy (*48*). For example, a vast assortment of moiré superlattices with ultraclean interfaces has been epitaxially grown, such as graphene/*h*-BN (*49*), WS_2/WSe_2 (*50*), and MoS_2/WSe_2 (*51*). Typically, high–symmetry-stacking configurations with the smallest formation energy are preferably deposited during the vdW epitaxial growth (*48*, *50*). Therefore, it is very challenging to obtain moiré superlattices with on-demand twist angles, especially the metastable samples (e.g., magic-angle TBG) that show strong electron-electron correlation and new optoelectronic phenomena. Breaking the energy tendency and realizing the scalable fabrication of moiré superlattices with controllable twist angles and layer-by-layer epitaxy would greatly promote the progress of the field.

Dynamic manipulation

For the "tear-and-stack" technique or vdW epitaxial growth, the final moiré superlattice

[1]QTF Centre of Excellence, Department of Electronics and Nanoengineering, Aalto University, Tietotie 3, FI-02150 Espoo, Finland. [2]Beijing National Laboratory for Condensed Matter Physics; Key Laboratory for Nanoscale Physics and Devices, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China. [3]School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100190, China. [4]Institute of Experimental Physics, Faculty of Physics, University of Warsaw, ul. Pasteura 5, 02-093 Warsaw, Poland. [5]Songshan-Lake Materials Laboratory, Dongguan, Guangdong Province 523808, China. [6]Department of Physics, University of California at Berkeley, Berkeley, CA, USA. [7]Institute of Photonics and Photon Technology, Northwest University, Xi'an 710069, China.
*Corresponding author. Email: zhipei.sun@aalto.fi (Z.S.); luojun.du@iphy.ac.cn (L.D.); gyzhang@iphy.ac.cn (G.Z.)

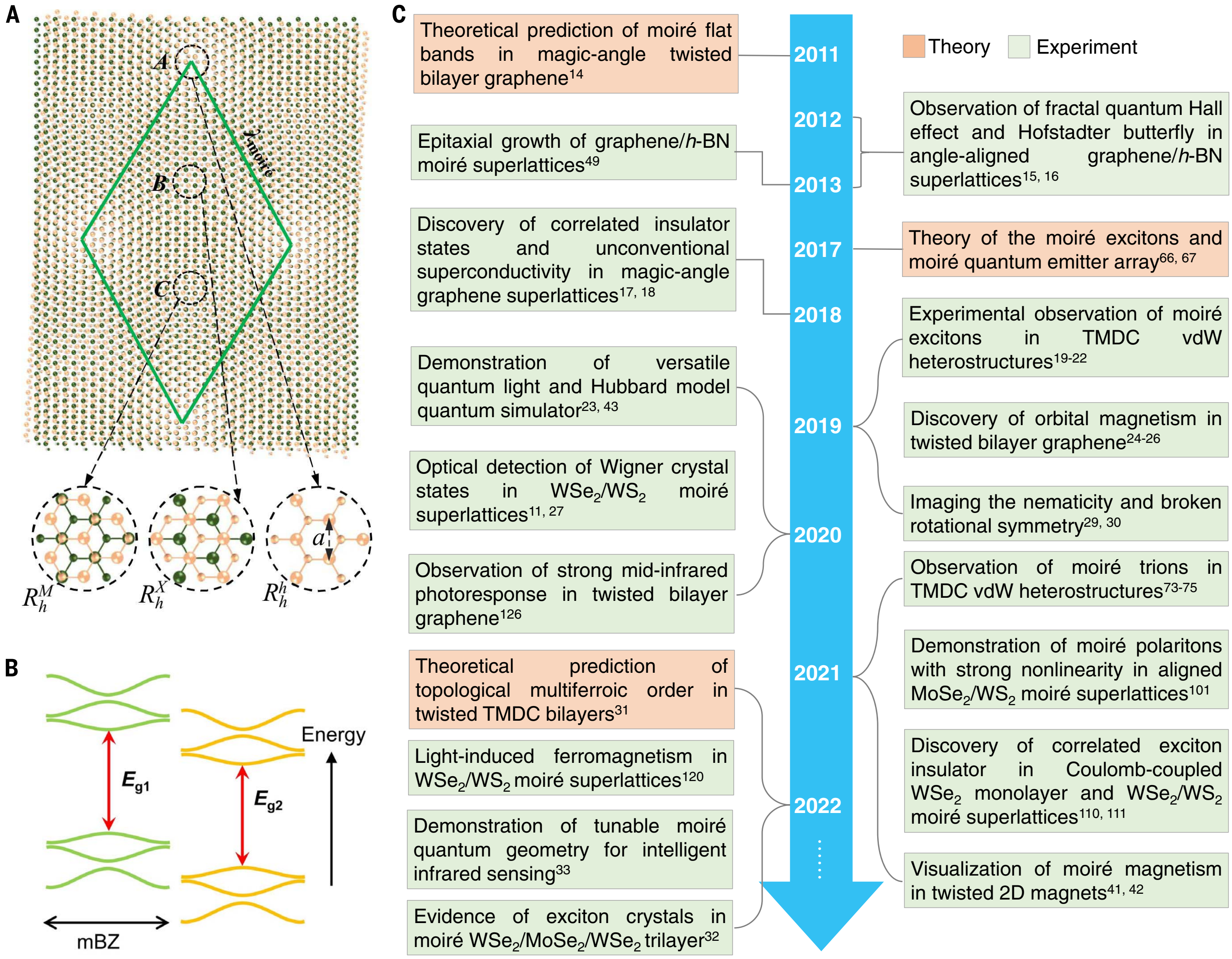


Fig. 1. Moiré superlattices and timeline of key developments in the study of moiré physics. (**A**) Top: Illustration of the moiré superlattice formed by a nearly 0°-angle–aligned TMDC heterobilayer. The moiré unit cell is outlined in the green diamond. Bottom: Close-ups of the three high-symmetry points (A, B, and C). Atomic registry R_h^{μ} denotes an *R*-type stacking, with the μ site of the top layer vertically aligned with the hexagon center (*h*) of the bottom layer. (**B**) Schematic of the flat moiré bands in mini Brillouin-zone (mBZ) for moiré heterostructures with type II band alignment. E_{g1} and E_{g2} stand for the bandgap of the two constituent TMDC layers. (**C**) The timeline of selected key developments in the study of moiré physics. Over the past few years, the field of moiré superlattices has grown substantially with a series of new discoveries, such as orbital magnetism, Wigner crystal states, electronic nematic phase, tunable spin-polarized correlated states, simulation of Hubbard model physics, continuous Mott transition, quantum anomalous Hall effect, moiré excitons/polaritons, strong mid-infrared photoresponse, light-induced ferromagnetism, and highly tunable bulk photogalvanic effects. [Adapted with permission from (A) (*66*)]

usually has a fixed twist angle. Therefore, it is often challenging to characterize the twist-angle–dependent behavior. Owing to the superlubricity of vdW interfaces, moiré superlattices with dynamically controllable twist angles can be obtained through mechanical, optical, thermal, and magnetic manipulation (*52*–*56*). This enables the feasibility to uncover the intrinsic twist-angle–dependent properties. For example, with the atomic force microscope tip manipulation technique, the twist-angle driven evolution of electronic, optical, and mechanical responses in graphene/*h*-BN moiré superlattices has recently been revealed (*53*).

Visualization of moiré superlattices

Direct visualization of moiré superlattices is fundamentally valuable for the comprehensive understanding and control over emergent moiré physics. In principle, atomic-resolution techniques, such as scanning tunnelling microscopy (STM) and transmission electron microscopy (TEM), offer the opportunity to directly image various moiré superlattices [e.g., magic-angle TBG (*29*), twisted bilayer WSe_2 (*57*), and aligned WS_2/WSe_2 (*20*, *50*)]. Specifically, STM can also probe the magnitude of moiré potential in situ, offering key insight for understanding the moiré physics. Although STM and TEM are powerful in the visualization of moiré superlattices, the special sample preparation and complex operating environment (e.g., ultrahigh vacuum and cryogenic temperature) are not conducive to establishing a clear correlation between the direct imaging of moiré structures and the emerging electrical and optical properties.

Because of the broken symmetry and sizable strain gradients at domain walls, moiré superlattices can show nonvanishing electromechanical responses and hence be imaged by piezoresponse force microscopy (PFM) (*58*). In particular, PFM provides a simple, universal

Table 1. Comparison between the visualization methods of moiré superlattices.

	Pros	Cons
Atomic-resolution techniques (such as TEM and STM)	1. Directly image the local displacement vector. 2. Direct visualization of the structural relaxation, disorder, and strain fields. 3. Directly visualize the magnitude of the moiré potential (for STM).	1. Require specialized sample preparation. 2. Require extreme conditions (e.g., ultra-high vacuum and low temperature). 3. Incompatible with most device fabrication techniques and optical/transport measurements.
AFM-based techniques (such as PFM, SKPM and conductive AFM)	1. Simple, room-temperature, ambient method. 2. Universal applicability. 3. Provide the key information of flexoelectricity (for PFM), surface potential (for SKPM), and interlayer hybridization (for conductive AFM).	1. Require direct contact with the active area and is incompatible with encapsulated samples. 2. Cannot image the local displacement vector. 3. Require conducting substrates (for conductive AFM).
Near-field techniques (such as SNOM, sMIM, and tip-enhanced Raman)	1. Ambient conditions. 2. Compatible with functional electronic devices.	Cannot image the local displacement vector.
Secondary electron imaging technique	1. Compatible with encapsulated samples and optical/transport measurements. 2. Universal applicability.	1. Cannot image the local displacement vector. 2. Cannot give the local stacking order.

technique to visualize various moiré superlattices under room-temperature and atmospheric conditions (58). However, PFM requires direct contact with the active area and thus prevents the use of h-BN encapsulation and top gates, which are typically required for most optical and transport measurements. Recent studies demonstrate that channeling modulated secondary electron imaging technique can be used to visualize the fully h-BN encapsulated, dual-gated moiré superlattice devices (59). With the help of this advanced technique, the local moiré structures and the exotic photonic properties have been successfully correlated with each other (59), deepening our understanding. Additionally, moiré superlattices can also be visualized by conductive atomic force microscopy (AFM) (60), scanning near-field optical microscopy (SNOM) (61), scanning microwave impedance microscopy (sMIM) (62), scanning Kelvin probe microscopy (SKPM) (63), and hyperspectral Raman imaging (64). Table 1 briefly compares the advantages and disadvantages of different visualization techniques of moiré superlattices.

Moiré photonics

Moiré superlattices with a long-wavelength periodic potential landscape offer opportunities to engineer the elementary excitations and light-matter interaction for fascinating photonic phenomena. In this section, we review the state-of-the-art moiré photonics, which mainly covers moiré neutral and charged excitons, resonantly hybridized excitons, moiré polaritons, moiré phonons, and reconstructed collective excitations.

Moiré neutral and charged excitons

The exciton, a Coulomb-bound electron-hole pair, dominates the optical properties of a material and hence underlies the development of emerging photonic and optoelectronic applications (65). Recent studies demonstrate that the moiré superlattices open up new avenues for modulating excitonic quasiparticles in both real and momentum spaces, resulting in quantum-dot–like moiré excitons and Bragg-umklapp moiré excitons, respectively (12).

Quantum-dot–like moiré excitons

From the perspective of real space, moiré superlattices impose a spatially periodic potential landscape on the excitons (66, 67). Therefore, excitons would likely be trapped at the moiré potential minima (Fig. 2A), forming a lattice of quantum-dot–like moiré excitons, a solid-state analog of a bosonic quantum gas in an optical lattice that is of interest for quantum photonics (19, 66). Recently, moiré-trapped quantum-dot-like interlayer excitons (IXs), evidenced as a series of discrete photoluminescence (PL) emission lines with ultranarrow linewidths (e.g., ~100 μeV), have been demonstrated in $MoSe_2/WSe_2$ heterostructures (Fig. 2B) (19, 68, 69). In particular, the moiré potential minima are located at high-symmetry sites that maintain threefold rotational symmetry (19, 66, 69). Consequently, moiré-trapped excitons can inherit valley-contrasting properties and exhibit strong circular polarization (Fig. 2B), a feature that distinguishes them from quantum emitters localized by random extrinsic potential (19, 70). In addition, owing to the quantum-dot–like confinement, moiré-trapped excitons could provide a platform for single-photon sources and quantum photonics, as demonstrated by the photon correlation experiments (Fig. 2D) (23).

Notably, quantum-dot–like moiré IXs show multiple discrete narrow PL lines at low excitation powers but transform into a broad emission peak as the excitation power increases (Fig. 2C) (19, 69). This feature can be understood as follows: At low excitation powers (e.g., a few nanowatts), the number of the photon-excited IXs are around two orders of magnitude smaller than the number of moiré sites, and thus are trapped at different moiré sites (Fig. 2C), because of the dipolar exciton-exciton interactions (69). Absent disorder, only one PL line is expected, as the potential minima in different moiré supercells would be degenerate. However, the ubiquity of inhomogeneity and disorder breaks the degeneracy, leading to spectral isolation of moiré-trapped IXs and a series of PL emission lines (19, 69, 71). As the excitation power increases, more IXs are excited, and thus more moiré trapping sites are populated, resulting in an ensemble band of moiré-trapped IXs with a broad linewidth (Fig. 2C) (69). When the exciton density is high enough, moiré traps would be filled up, and excitons approach the regime of free-particle behavior (19, 72, 73).

Moiré-trapped trions

Upon doping of moiré superlattices, moiré-trapped neutral excitons would interact with the excess charge carriers that are also localized at the moiré sites, forming the moiré charged excitons (i.e., moiré-trapped trions) (74, 75). Recently, moiré-trapped IX trions have been demonstrated in $MoSe_2/WSe_2$ heterostructures (69, 73–75). In these experiments, new sets of quantum-dot–like IX trion emission lines with ultranarrow linewidths appear about 7 meV below the moiré-trapped neutral excitons upon electrostatic doping (Fig. 2E) (74, 75).

Moiré-trapped trions, because of their charged nature, can couple with the electrons or holes in nearest-neighbor moiré sites via long-range Coulomb interactions. And because the charge

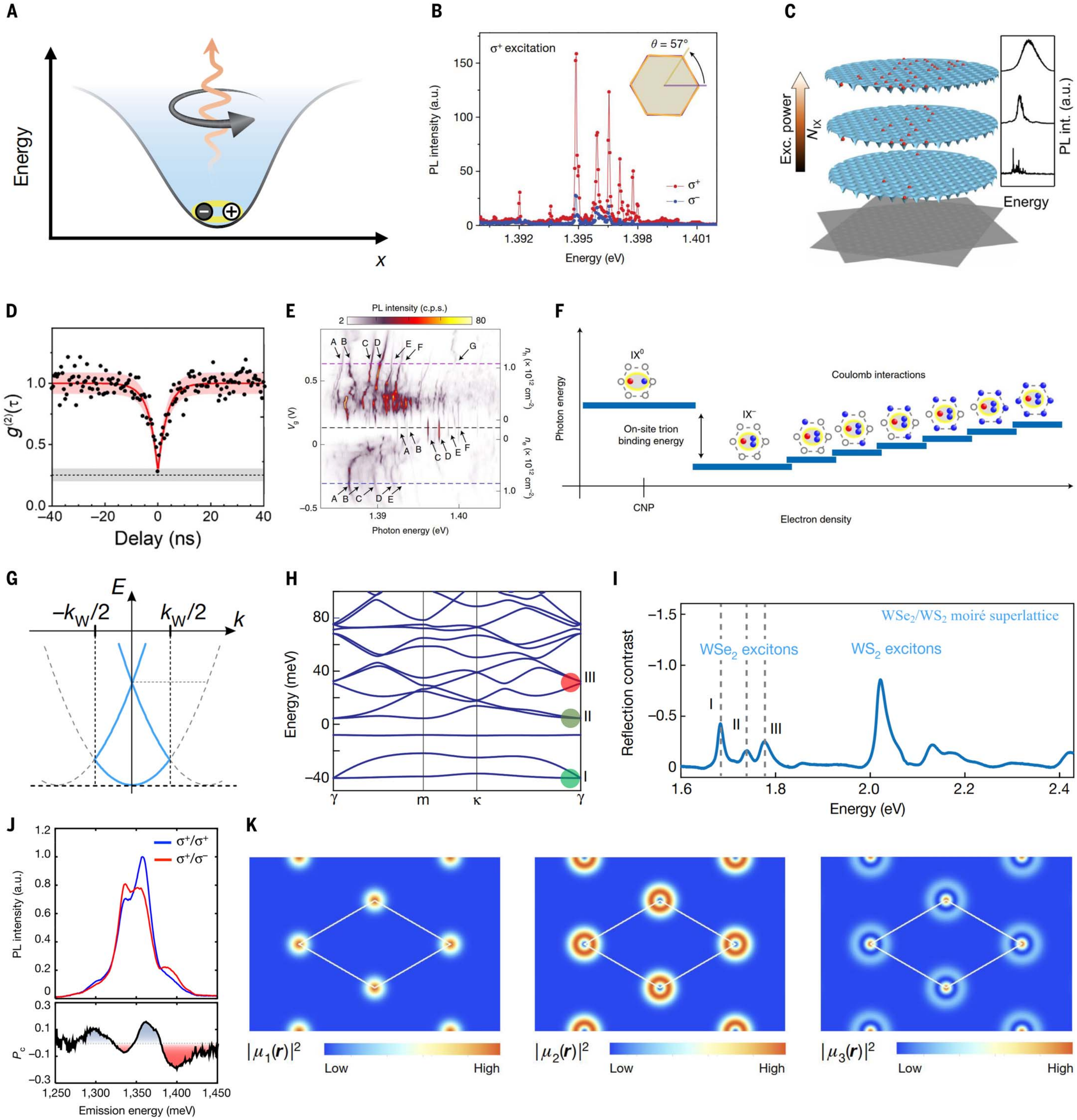


Fig. 2. Moiré excitons and trions. (**A**) Representation of an exciton trapped at the moiré potential minima. (**B**) Circularly polarized PL spectra of moiré-trapped IXs of 57° $MoSe_2/WSe_2$ heterostructures. (**C**) Schematic of the occupation of moiré sites by IXs with increasing excitation power. (**D**) Second-order photon correlation statistics of moiré-trapped IXs. The antibunching indicates the single-photon nature. (**E**) PL map as a function of doping. At electron or hole doping, moiré-trapped trions with lower energy emerge. The prominent IX emissions are labeled as A to G. (**F**) Schematic of the moiré-trapped IX energy against electron density. Successive filling of nearest-neighbor moiré sites leads to a Coulomb staircase. (**G**) Sketch depicting the exciton band folding by umklapp scattering off moiré potential. (**H** and **I**) Calculated WSe_2 A exciton dispersion of WS_2/WSe_2 moiré superlattices, with three optically allowed transitions denoted I to III (H), which are observed by reflectance contrast spectra (I). (**J**) Helicity-resolved PL spectra of 1° $MoSe_2/WSe_2$ heterobilayers, with alternating circular polarization. (**K**) Real-space map of the center-of-mass wave functions for the first three umklapp moiré IXs. [Adapted with permission from (A and B) (*19*), (C) (*69*), (D) (*23*), (E and F) (*74*), (H and I) (*20*), (J and K) (*21*)]

carriers are also moiré-trapped, the six nearest-neighbor moiré sites are filled sequentially (Fig. 2F) (74). This creates a discrete repulsive Coulomb interaction between moiré-trapped trions and the charge carriers in nearest-neighbor moiré sites. Consequently, moiré-trapped trions show a staircase-like blueshift of energy (Fig. 2F), providing a noninvasive, local optical technique to characterize the Coulomb interactions and local disorder in moiré superlattices (74). In addition, moiré-trapped trions exhibit distinctive valley properties. For example, the valley polarization can be switched from large cocircular polarization under electron doping to negligible cross-circular polarization under hole doping (74, 75), and the valley polarization time can reach ~1 μs, three orders of magnitude longer than that of conventional trions in monolayer TMDCs (75, 76). This opens up possibilities for opto-valleytronic applications and emerging nanophotonic devices, such as spin-valley optical transistors and chiral light-emitting diodes.

Bragg-umklapp moiré excitons

From the perspective of momentum-space, Bragg-umklapp scattering off the long-period moiré potential can fold the dark exciton states with momentum $k = k_M$ (where k_M is the moiré reciprocal lattice vectors) into the center of the mBZ, resulting in the formation of new excitonic minibands with optical activity (Fig. 2G) (12, 22, 67, 77–80). Figure 2H illustrates the WSe_2 A exciton dispersion of aligned WSe_2/WS_2 superlattices and clearly shows three optically active excitons at the mBZ center (I, II, and III) (20, 67). Experimentally, Bragg-umklapp intralayer moiré excitons have been demonstrated by both absorption spectroscopy in WSe_2/WS_2 heterobilayers (Fig. 2I) (20, 43) and PL measurements in $MoSe_2/MoS_2$ heterostructures (81). Notably, the cutting-edge exciton band structure calculations reveal that different intralayer moiré excitons exhibit distinct hallmarks (82). In WSe_2/WS_2 moiré superlattices, the lowest-energy moiré resonance I is characterized by a Wannier-type exciton with the tightly correlated electron and hole, whereas the highest-energy moiré exciton peak III is described by an intralayer charge-transfer exciton with spatially separated electron and hole (82). Such distinct characters provide unparalleled insight into the unusual doping responses and magneto-optical responses of moiré excitons (20, 82).

For IX species, Bragg-umklapp moiré exciton transitions have also been uncovered by PL spectra in $MoSe_2/WSe_2$ superlattices (Fig. 2J) (21, 83). Owing to the strongly modified exciton center-of-mass wave function in real space, Bragg-umklapp moiré IX resonances exhibit alternating positive and negative circular polarization. Figure 2K shows the real-space distribution of center-of-mass wave functions for the first three Bragg-umklapp moiré IX resonances of $MoSe_2/WSe_2$ heterostructures, which correspond to s-, p- and d-wave symmetry, respectively (21, 79). Because of an additional angular momentum $\hbar$ ($2\hbar$), Bragg-umklapp moiré IX with the p-waveform (d-waveform) has the opposite (same) optical selection rule as the first s-wave symmetry one, leading to alternating co- and cross-circularly polarized PL (21).

In the presence of itinerant electrons or holes, Bragg-umklapp moiré trions can appear in moiré superlattices (84). Besides excitons and trions, Bragg-umklapp scattering off the moiré potential can also activate other elementary excitations, such as phonons, magnons, and polaritons. For example, Bragg-umklapp moiré phonons have been uncovered in TBG (85, 86), graphene/*h*-BN superlattices (86, 87), and twisted bilayer TMDCs (64, 88).

We remark that quantum-dot–like moiré excitons from the real-space perspective and Bragg-umklapp moiré excitons from the momentum-space perspective should be unified in principle. For instance, theoretical calculations show that the first few Bragg-umklapp moiré excitons with low energy are localized at the moiré potential minima and thus should also belong to the quantum-dot–like moiré states (21, 78, 79). Alternatively, quantum-dot–like moiré excitons observed under ultralow excitation power should correspond to the Bragg-umklapp moiré excitons with the lowest energy. However, quantum-dot–like moiré excitons and Bragg-umklapp moiré excitons demonstrated by current experimental measurements seem to exhibit distinctly different behaviors. For example, the linewidths of Bragg-umklapp moiré excitons are more than two orders of magnitude wider than those of quantum-dot–like moiré excitons. These discrepancies may stem from the ubiquity of disorder, strain, or twist-angle inhomogeneity in actual real-world moiré superlattices and deserve further studies to establish a unified understanding.

Resonantly hybridized excitons

Usually, the effects of moiré superlattices on excitons are regarded as a harmonic perturbation potential, then the intralayer excitons and IXs are discussed separately as independent objects (such as the previous section) (66, 67). Such a view holds true for moiré superlattices with a large band-edge offset. On the contrary, when the conduction and/or valence band edges of the constituent layers are aligned in both momentum and energy, the resonant interlayer hybridization can strongly enhance the moiré superlattice effects and mix the intralayer excitons and IXs through spin-conserving interlayer hopping, resulting in the resonantly hybridized excitons (12, 22, 89). Resonantly hybridized excitons combine the merits of both intralayer excitons (e.g., large oscillator strength) and IXs (e.g., appreciable out-of-plane dipole moment), and thus represent an advantageous scenario for nanophotonic and optoelectronic applications (90, 91). Resonantly hybridized excitons have been demonstrated in TMDC homobilayer moiré superlattices with intrinsically aligned band edges (40, 89, 92–94), $MoSe_2/WS_2$ moiré heterostructures with nearly degenerate conduction-band edges (Fig. 3A, right panel) (22, 89–91), and WSe_2/WS_2 moiré superlattices where the upper valence band of WS_2 and the lower valence band of WSe_2 are closely aligned (Fig. 3A, left panel) (91). Furthermore, according to the available electron band parameters in the literature (95), we envision that resonantly hybridized excitons can also appear in MoS_2/WS_2, $MoSe_2/WSe_2$, $MoSe_2/MoTe_2$, $MoTe_2/WS_2$, and $MoTe_2/WSe_2$ moiré superlattices.

Resonantly hybridized excitons display a sharp energy modulation against the twist angle (Fig. 3B) (22, 89, 90) and manifest energy-level anticrossing under the applied electric fields (Fig. 3C) (89, 91, 92). It is instructive to understand these features using a two-level coupled-oscillator model with Hamiltonian $\begin{pmatrix} \varepsilon_X & \Omega \\ \Omega & \varepsilon_{IX} \end{pmatrix}$, where the off-diagonal element Ω denotes the coupling constant and the diagonal elements ε_X (ε_{IX}) describe the energy of uncoupled intralayer exciton (IX). Because the band minimum of IX is located at the corners of mBZ, IXs that can hybridize with optically bright intralayer excitons should have energy $\varepsilon_{IX} = \varepsilon_{min} + \frac{\hbar^2 q^2}{2M_{IX}}$, where M_{IX} is the total mass of IX, ε_{min} is the IX energy at band minimum, and q denotes the momentum of mBZ corners (22, 79, 89, 90). Given that q increases linearly with twist angle, ε_{IX} and hence resonantly hybridized excitons exhibit a sharp energy modulation (Fig. 3B) (22, 89, 90). Meanwhile, IX shows a linear energy shift with applied electric fields due to the quantum-confined Stark effect, leading to the electrically tunable resonantly hybridized excitons (91, 96, 97). When the energy of IX is tuned to resonate with ε_X, resonantly hybridized excitons would show avoided crossing of the energy level, as demonstrated by absorption and PL spectra (Fig. 3C) (91, 92). For WSe_2/WS_2 moiré superlattices, the extracted coupling constants Ω can reach ~40 meV, much larger than the exciton linewidths (~10 meV). This indicates that the system is in the exciting strong-coupling regime and provides opportunities for engineering various emerging many-body states, such as Bose-Einstein condensation and polariton superfluidity (91, 98).

Moiré polaritons

Polaritons, a hybrid quasiparticle excitation of part-light and part-matter, enable the fine

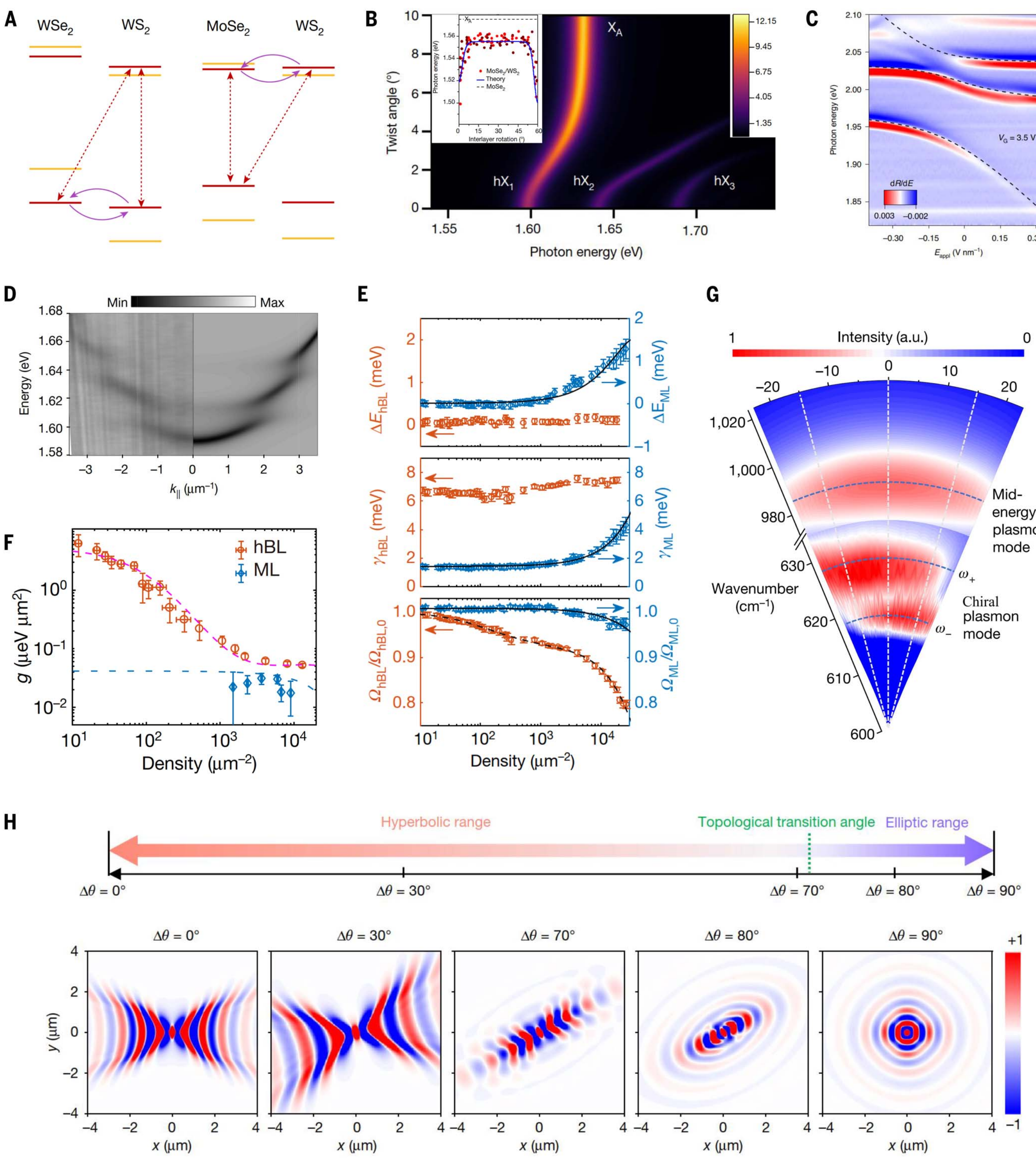


Fig. 3. Resonantly hybridized excitons and moiré polaritons. (**A**) Schematic diagram of resonantly hybridized excitons resulting from the interlayer hole (right) and electron hopping (left). (**B**) Calculated absorption spectrum of resonantly hybridized exciton versus twist angle in $MoSe_2/WS_2$ moiré superlattices, which is demonstrated by PL measurements (inset). (**C**) Derivative of the reflectance spectrum against the electric field in 60°-aligned WS_2/WSe_2 moiré superlattices. (**D**) Angle-resolved reflectance spectra, demonstrating strong coupling between resonantly hybridized excitons and cavity photons. (**E**) Carrier density–dependent shift of exciton energy (upper), half-linewidth (middle), and normalized coupling strength (lower) for lower polaritons (LPs) of $MoSe_2/WS_2$ moiré superlattices (denoted as hBL) and monolayer $MoSe_2$ (denoted as ML). (**F**) Nonlinear coefficient, g, as a function of carrier density for moiré hBL LPs (red) and ML LPs (blue). (**G**) Extinction mapping of chiral Berry plasmons against wave number and polarization detection angle, demonstrating the magnetic-field–free Faraday effect. (**H**) Topological transitions triggered by the twist angle from open (hyperbolic) to closed (elliptical) dispersion contours in bilayers of α-MoO_3. [Adapted with permission from (B) (22), (C) (91), (D to F) (101), (G) (104), (H) (108)]

control of light at the nanoscale across a broad spectral range, promising the prospect of new technological applications in nanophotonics and optoelectronics (*99*, *100*). Recent studies show that moiré superlattices offer an opportunity for polariton engineering and photonic quantum technologies. Here, we provide an overview of the state-of-the-art moiré exciton-, plasmon-, and phonon-polaritons.

Moiré exciton-polaritons

Because of appreciable oscillator strengths and in-built electric dipoles, resonantly hybridized moiré excitons hold great promise for polaritons with quantum nonlinearity and cooperative effects. $MoSe_2/WS_2$ moiré superlattices with resonantly hybridized moiré excitons are integrated into a microcavity (*101*). In place of the resonantly hybridized moiré excitons or the bare cavity photons, three new light-matter hybrid modes with anticrossing of the dispersion curve emerge (Fig. 3D) (*101*). This demonstrates the moiré exciton-polaritons and the effective coupling between resonantly hybridized excitons and cavity photons. Moreover, the coupling strengths Ω (~10 meV) of moiré exciton-polaritons satisfy $\Omega > \frac{(\gamma_X+\gamma_p)}{2}$ (γ_X/γ_p are the linewidths of the resonantly hybridized excitons and cavity mode), indicating the strongly coupled regime of light–matter interactions (*98*, *101*).

Moiré exciton-polaritons exhibit radically different excitation density responses than polaritons formed with the constituent monolayer excitons (Fig. 3E) (*101*). For example, with increasing excitation density, the moiré exciton-polaritons (monolayer exciton-polaritons) show a strong drop in coupling strength by up to 20% (negligible change in coupling strength), but no measurable exciton energy shift or line broadening (considerable blueshift and line broadening). This suggests the strong exciton blockade and suppressed excitation-induced dephasing for moiré exciton-polaritons, and indicates that the resonantly hybridized exciton involved in the polariton formation should be a moiré-trapped 0D-like exciton (*101*). Benefiting from the moiré quantum confinement effect, moiré exciton-polaritons can show a nonlinearity up to ~6 $\mu eV \cdot \mu m^2$, more than two orders of magnitude larger than typical monolayer TMDC exciton-polaritons (Fig. 3F) (*101*). Moiré exciton-polaritons combine strong quantum nonlinearity and large coupling strength, opening a huge swath of possibilities for quantum photonics and optoelectronics.

Moiré plasmon-polaritons

Plasmon-polaritons are coupled excitations of photons and charge carriers that offer the possibility of manipulating and guiding light at subwavelength scales (*99*, *100*). The ability to control plasmon-polaritons is central to modern information and communication technologies, such as subdiffraction imaging, optical sensing, infrared detectors, light-harvesting devices, and electro-optical modulators (*99*, *100*). Moiré superlattices provide a powerful platform for engineering new plasmon-polaritons because of the interlayer quantum coupling and radically altered electronic structures (e.g., the emergence of flat bands). For example, owing to additive contribution from the interband transition between superlattice minibands, plasmon-polaritons in graphene/*h*-BN moiré superlattices show enhanced scattering amplitudes (*102*). For TBG moiré superlattices, an exotic interband plasmon-polariton mode originating from the optical transitions between flat minibands is observed, which is radically different from the conventional intraband plasmon in monolayer graphene (*103*). By using nanostructured surface plasmon resonators, two new plasmon modes are identified in TBG moiré superlattices: a chiral Berry plasmon mode from uncompensated Berry flux, and a slow plasmon mode from interband transitions between nested subbands (*104*). In particular, because of the nontrivial Berry curvature, the chiral plasmon mode can show the magnetic-field–free Faraday effect with a polarization rotation of ~15° (Fig. 3G) (*104*). Such a large Faraday rotation angle is comparable to the conventional Faraday effect accessible by a static magnetic field of 7 T, highlighting the great potential of magnetic-field–free Faraday rotators (*104*). The slow plasmon mode is within the highly sought-after mid-wave infrared spectral window and would open up opportunities for new photonic devices, such as infrared devices, without fundamental damping and phonon scattering processes (*104*). It is noteworthy that numerous plasmon-polaritons, such as intrinsically undamped plasmon modes (*105*), quasi-flat plasmonic bands (*106*), and plasmonic Dirac cone (*107*), have been predicted in moiré superlattices and deserve further exploration.

Moiré phonon-polaritons

Phonon-polaritons, collective oscillations that result from the hybridization of photons and optical phonons in polar crystals, can bestow a wealth of valuable applications, such as nanoscale light propagation, ultramicroscopy, and coherent thermal emission (*99*, *100*). Experimental studies have demonstrated that in twisted stacks of vdW crystal α-MoO_3, the interlayer twist angle enables fine control of the dispersion of phonon-polaritons and therefore triggers a topological transition from hyperbolic (open) to elliptical (closed) wavefront geometries (Fig. 3H) (*108*, *109*). Such a transition is determined by a topological quantity and becomes resilient to perturbations, impurities, or disorder. When the topological number is two (four), the dispersion of phonon-polaritons is hyperbolic (elliptical) (*108*, *109*). At the twist angle at which the topological transition occurs (dubbed as photonic magic angles), the dispersion becomes flattened, supporting highly collimated, directive, and diffractionless canalized phonon-polaritons for various applications, such as quantum nanophotonics, superlens, nanoimaging, biosensing, and radiative energy transfer (*108*).

Photonics of moiré correlated electronic states

Moiré superlattices, with enhanced electron-electron correlations in the flat minibands, provide opportunities for a large portfolio of correlated electronic states, such as Mott insulators, Wigner crystal states, charge density waves, charge-transfer insulators, and exciton insulators (*11*, *27*, *28*, *110*–*113*). For example, nearly two dozen correlated insulating states have been unveiled at the commensurate filling fractions of WSe_2/WS_2 moiré superlattices via an optical sensing technique (Fig. 4A) (*11*). Certainly, these moiré correlated electronic states would have profound effects on the elementary optical excitations, resulting in a number of photonic responses.

First, owing to the gapped nature of moiré correlated electronic states, free carrier screening of the exciton interactions is largely quenched. This would result in the enhancement of oscillator strength and the blueshift of exciton resonance energy for photonics (Fig. 4B) (*30*, *114*, *115*). Second, moiré correlated electronic states at commensurate fractional fillings belong to charge-ordered states, e.g., Wigner crystals and charge density waves. Consequently, there is an additional periodic potential for collective excitations, which can fold dark states back to the light cone by Bragg-umklapp scattering and thus brighten new optical transitions (Fig. 4C) (*80*, *116*). Third, for moiré correlated electronic states at some specific fillings (e.g., $v = \pm\frac{1}{2}, \pm\frac{3}{5}$), electrons are arranged in alternating lines with a preferential orientation along the high-symmetry axes of the moiré superlattice (*11*, *30*, *112*). Such stripe nature of these moiré correlated electronic states spontaneously breaks the threefold rotational symmetry, enabling the anisotropic optical responses (*30*, *117*). Last but not least, moiré correlated electronic states can facilitate a wide range of magneto-optical phenomena, such as strongly enhanced valley Zeeman splitting, magnetic circular dichroism, Faraday effect, and Kerr rotation (*43*, *118*, *119*). For example, a nearly two-orders-of-magnitude–enhanced exciton valley Landé *g*-factor has been observed for moiré correlated electronic states at $v = -1$ (Fig. 4D), opening up possibilities for opto-valleytronic applications (*43*, *118*). By optically tuning the spin exchange interactions between moiré-trapped holes, ferromagnetic order emerges at $v = -\frac{1}{3}$ Wigner state of WSe_2/WS_2 moiré superlattice, giving

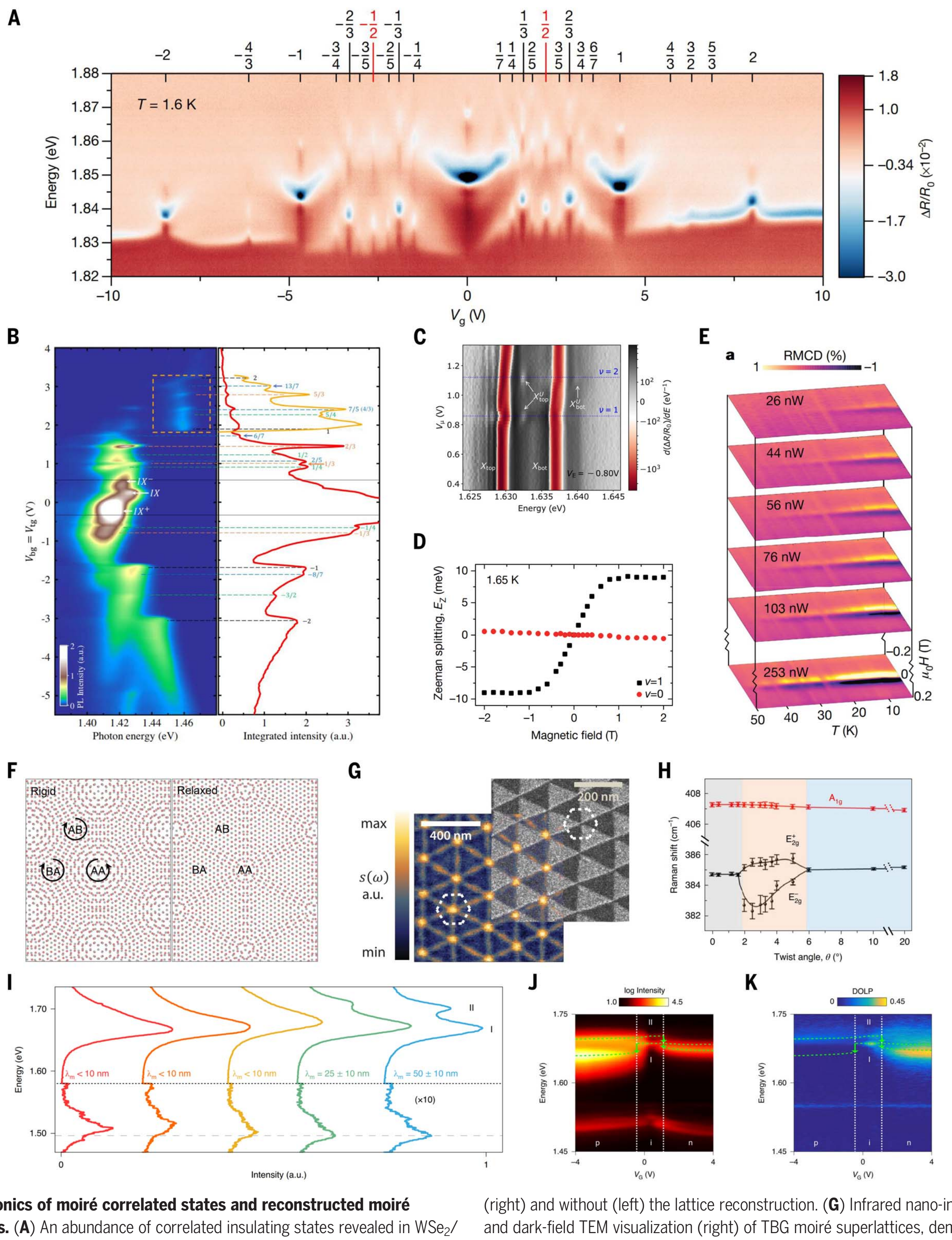


Fig. 4. Photonics of moiré correlated states and reconstructed moiré superlattices. (**A**) An abundance of correlated insulating states revealed in WSe_2/WS_2 moiré superlattices by optical sensing technique with 2s exciton resonance. (**B**) Gate-dependent PL map (left) and integrated intensity (right) for the IX states in a WS_2/WSe_2 moiré superlattice. (**C**) Dependence of chemical potential of differential reflectance differentiated with respect to energy. Umklapp exciton resonances XU top and XU bot can be clearly observed at $v = 1$ and $v = 2$. (**D**) Exciton valley Zeeman splitting as a function of the magnetic field at zero filling (red circles) and half-filling (black squares). (**E**) Reflective magnetic circular dichroism (RMCD) as a function of magnetic field and temperature at selected excitation power, indicating light-induced ferromagnetism. (**F**) Schematic cartoon of TBG moiré superlattices with (right) and without (left) the lattice reconstruction. (**G**) Infrared nano-imaging (left) and dark-field TEM visualization (right) of TBG moiré superlattices, demonstrating a natural plasmon photonic crystal. (**H**) The energies of intralayer phonon modes versus twist angle. Clear peak splitting of E_{2g} mode can be observed at 2° ≤ θ < 6°, providing information on reconstructed collective excitations. (**I**) PL spectra of twisted bilayer WSe_2 collected from the locations with different moiré wavelengths. A higher-energy exciton peak (type II) emerges with increasing moiré wavelength. (**J** and **K**) Gate-dependent PL intensity (J) and linear polarization (DOLP) of reconstructed twisted WSe_2 bilayers (K), showing distinct gate tunability and valley coherence properties. [Adapted with permission from (A) (11), (B) (114), (C) (116), (D) (43), (E) (120), (F) (121), (G) (61), (H) (124), (I to K) (59)]

rise to the exotic magnetic circular dichroism (Fig. 4E) (120). Additionally, giant Kerr rotations, over an order of magnitude larger than those observed in typical materials, are predicted in moiré correlated electronic states (119). Nevertheless, coupling strongly correlated electronic states in moiré superlattices with light provides an exciting platform to create, probe, and manipulate strongly correlated states and new light-matter hybrids (e.g., photon-mediated superconductivity, ferroelectricity, or ferromagnetism; optically driven topological phenomena; highly correlated light-matter states; and optical spin Hall effect).

Reconstructed collective excitations

In the prior discussion, the influence of moiré superlattices on optical properties is based mainly on a rigid lattice model (Fig. 4F, left panel). By contrast, for moiré superlattices with a large period, the interlayer vdW interaction can compete with the intralayer lattice distortion and favor interlayer commensurability to minimize the misalignment energy, resulting in lattice reconstruction and thus breaking down the rigid lattice picture (60, 121, 122). For instance, a tessellated pattern of AB/BA triangular domains, separated by a network of narrow strain solitons, is formed in marginally twisted graphene and TMDC homobilayers (Fig. 4F, right panel) (60, 121, 122). Undoubtedly, the reconstructed moiré superlattices would cause renormalization of the electronic and optical properties. It was reported that the domain-wall solitons of marginally TBG are topologically protected chiral states and can strongly reflect the plasmon-polaritons (Fig. 4G) (61, 123). Consequently, a regular pattern of plasmon-polariton scatterers with a periodicity close to its wavelength is formed, transforming the TBG moiré superlattice into a natural nano-light photonic crystal (61). Crucially, such moiré photonic crystals are lithography-free and electrically tunable, representing an advantageous scenario for optical and optoelectronic telecommunications, as well as controllable quantum optical circuits.

In addition, the domain-wall solitons and soliton interceptions have considerable strain, facilitating strain-engineered photonics. For example, the emergence of strain in reconstructed graphene or MoS_2 moiré superlattices distorts the hexagonal unit cell, leading to the splitting of the doubly degenerate E_{2g} phonon mode (Fig. 4H) (124, 125). The reconstructed moiré superlattices also have profound impacts on symmetry-breaking photonics. For near-0°-twist-angle TMDC homobilayers, atomic reconstruction leads to the formation of energetically favorable AB and BA stacking. Owing to the simultaneous breaking of inversion and mirror symmetries, rhombohedral AB and BA stacking configurations show out-of-plane ferroelectric polarization, but with opposite signs. Therefore, the layer symmetry in reconstructed TMDC moiré superlattices is broken, enabling the emergence of two spatially alternating exciton species with distinct gate tunability and valley coherence properties (Fig. 4, I to K) (40, 59). In addition, reconstructed TMDC moiré superlattices open up new avenues to realize tunable exciton arrays and 1D localized exciton states with electrons and holes residing at opposite sides of the domain-wall solitons for applications such as quantum optoelectronics (59).

Moiré optoelectronics

Optoelectronics involves the study and applications of electronic devices that can generate, modulate, detect, interact with, or control light, including photodetectors, light-emitting diodes, photovoltaics, modulators, and so on. Moiré superlattices with combined intriguing electronic and optical properties point to new avenues for cutting-edge optoelectronic applications. In this section, we briefly review the recent developments in state-of-the-art moiré optoelectronics.

Strong mid- and far-infrared photoresponses

In general, moiré superlattices are expected to show strong light–matter interaction and large photoresponses. First, the formation of flat minibands with a large density of states can enhance the dynamical conductivity. Second, the emergence of superlattice bandgap and correlated insulating gap on the order of tens of millielectronvolts can lead to resonance interband transition and thus favor strong photoresponses in the long-sought mid- and far-infrared and even terahertz ranges. When the Fermi level lies in the superlattice bandgap (Fig. 5A), the interband transition between the top of the moiré Dirac band and the bottom of the empty band leads to the resonance-enhanced dynamical conductivity and hence strong mid-infrared photoresponses in TBG moiré superlattices (Fig. 5B) (126). A giant extrinsic photoresponsivity of ~26 mA W^{-1} at the mid-infrared wavelength of 12 μm is achieved in 1.81° TBG, which is almost a factor of 20 larger than the photoresponsivity of natural bilayer graphene (<1.5 mA W^{-1}) (126). As a result, TBG moiré superlattices provide an attractive material platform for the long-sought-after mid-infrared optoelectronic devices.

By using the optical transition between moiré flat minibands, strong far-infrared photocurrent responses have been seen in ABC trilayer graphene/h-BN moiré superlattices (Fig. 5C) (127). Owing to the electrically tunable moiré minibands, including both the bandwidth and bandgap, the photocurrent responses show high tunability over 80 meV in external electric fields (Fig. 5C, inset) (127). Upon doping the flat moiré miniband to form correlated insulating states, optical transition across the emerging Mott gap can largely enhance the dynamical conductivity and result in strong far-infrared photoresponses. For example, large photoresponses centered at ~18 meV are recently demonstrated at v = ±1/2 Hubbard-Mott insulating states of ABC trilayer graphene/h-BN moiré superlattices (Fig. 5D) (127, 128). We highlight that the strong far-infrared photocurrent responses induced by a correlated insulating gap in moiré superlattices can favor exotic optoelectronics and technological advances (e.g., detectors, sensors, and modulators).

Symmetry-breaking optoelectronics

Moiré superlattices with strong electron correlations usually display various symmetry-breaking [e.g., broken inversion symmetry, broken time-reversal symmetry, and broken U(1) gauge symmetry] and thus provide a powerful platform for symmetry-breaking optoelectronics, such as bulk photovoltaic effect (BPVE), opto-valley Hall effects, and chiral light-emitting diodes (3, 34, 129). For example, it was shown that because of moiré-induced symmetry breaking and quantum geometrical properties, strong BPVE at ~5 and 7.7 μm can occur in twisted double bilayer graphene moiré superlattices (Fig. 5E) (33). The moiré quantum geometry strongly depends on the displacement field and Fermi level, resulting in electrically tunable BPVE (e.g., the photovoltage amplitude and phase) in twisted double bilayer graphene superlattices. In addition, the mid-infrared BPVE in twisted double bilayer graphene is surprisingly strong and reaches ~3.7 V W^{-1} at 7.7 μm, much larger than previous demonstrations (33). We envision that the ultrastrong mid-infrared BPVE in moiré superlattices will stimulate next-generation applications in nonlinear optics and optoelectronics, such as intense terahertz sources, biological sensing, energy harvesting, and all-optical transistors.

The gate-tunable BPVE in twisted double bilayer graphene moiré superlattices, together with its dependence on incident optical polarization states, power, and wavelength, provides the possibility for the simultaneous decipherment of full-Stokes polarimetry and wavelength. Indeed, using the tunable mid-infrared BPVE in twisted double bilayer graphene as an encoder and a trained convolutional neural network as a decoder (Fig. 5F), a compact intelligent infrared sensor with merely a sub-wavelength footprint is achieved (33). After proper training of the convolutional neural network, the intelligent infrared sensing can identify the fingerprint of light, including the full Stokes parameters and the wavelength (Fig. 5G) (33), a pathway for future intelligent sensing technologies in an extremely compact, on-chip manner.

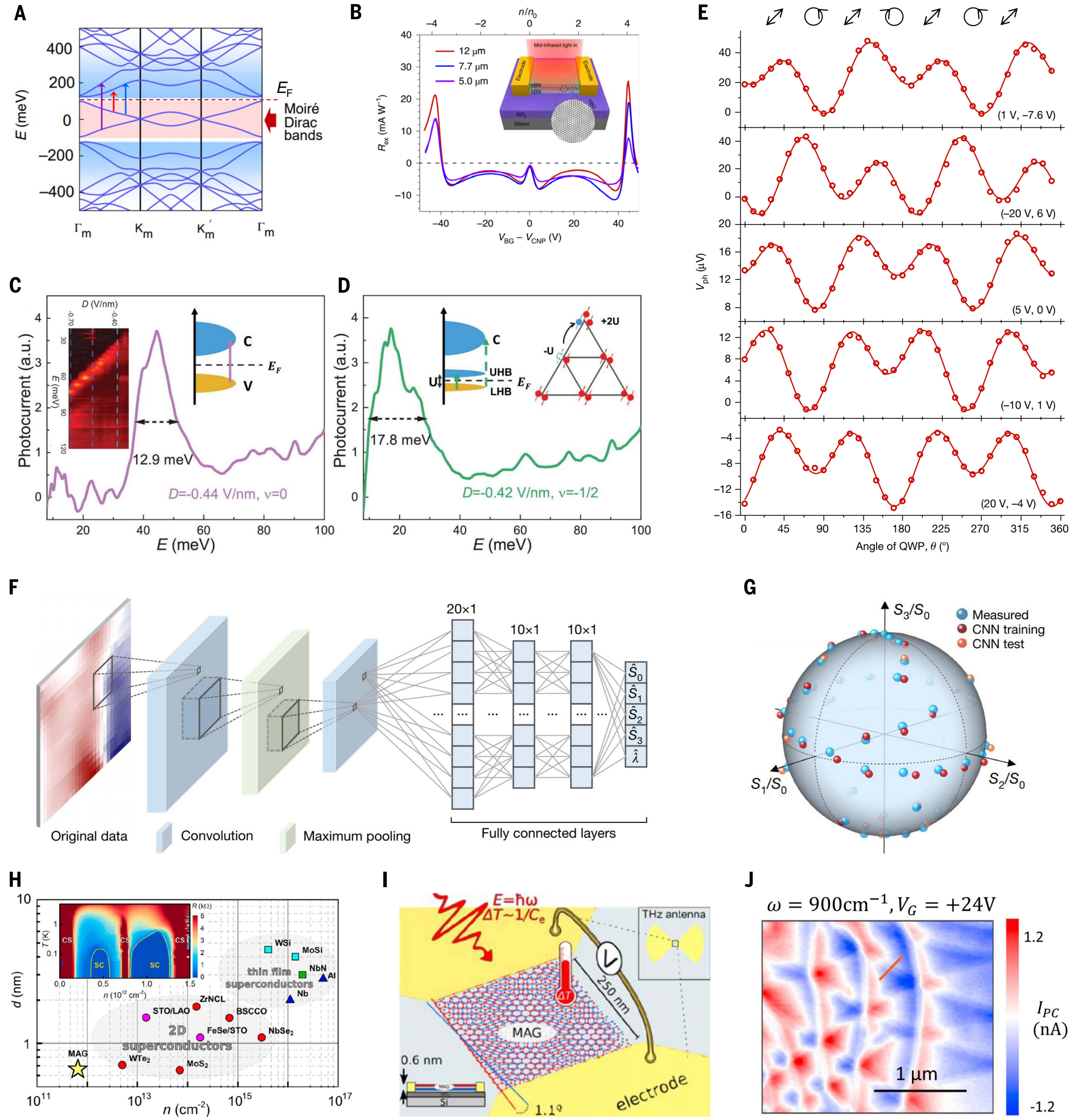


Fig. 5. Moiré optoelectronics. (**A**) Calculated band structure of 1.81° TBG, showing superlattice-induced bandgaps below and above moiré Dirac bands. (**B**) Photoresponses of TBG moiré superlattices at 5.0-, 7.7-, and 12-µm light illumination. Inset: Schematic diagram of the moiré superlattice device for photocurrent measurement. (**C** and **D**) The photocurrent spectra of ABC trilayer graphene/*h*-BN moiré superlattices at zero-filling (C) and half-filling of the flat valence band (D). Photocurrents in (C) and (D) are dominated by a sharp peak at ~45 meV and a broad peak at ~18 meV, respectively, corresponding to excitation between moiré conduction and valence minibands or to excitation across the emerging Mott gap, as illustrated by the inset. The inset on the right in (D) shows the optical transition across the Mott gap. (**E**) Photovoltage against the angle of a quarter-wave plate (QWP) at different gate voltage biases. (**F**) Schematic diagram of the convolutional neural network used for detection of the full-Stokes and wavelength. (**G**) Comparison of the polarization states obtained from convolutional neural network (CNN) outputs (red or orange spheres for training or test data) with those directly measured values (blue spheres), showing high accuracy. (**H**) Comparison of carrier density and thickness between superconductor magic-angle TBG and 2D, thin-film superconductors below 10 nm. Inset: Phase diagram of magic-angle TBG. (**I**) Schematic cartoon of magic-angle TBG superconducting single-photon nanocalorimeter. (**J**) Nano-imaging photocurrent of a TBG moiré superlattice. [Adapted with permission from (A and B) (126), (C and D) (127), (E to G) (33), (H and I) (134), (J) (135)]

Apart from the mid-infrared BPVE in twisted double bilayer graphene (*33*), it was also shown that the symmetry of TBG close to the second magic angle (~0.6°) is reduced to point group C_1, breaking both the inversion and rotational symmetries, and enabling the exotic BPVE in the terahertz region (*130*). It is expected that highly tunable BPVE can also emerge in other moiré superlattices and is worthwhile to explore, especially for rotationally aligned graphene/*h*-BN, hetero-, and homo-TMDC bilayers, in which electron correlation–triggered symmetry breaking and giant second-order nonlinear responses have recently been uncovered by electrical transport (*131*, *132*).

Single-photon detection

The detection of light at the single-photon level is crucial for a host of theoretical, experimental, and technological advances, such as quantum information processing, quantum sensing, quantum key distribution, and radio astronomy (*70*). By using heat-induced breaking of the superconducting states in nanostructured superconductors, advanced single-photon detectors with operation wavelengths in the range between visible and near-infrared have been well developed and even commercialized. However, because of high electron density and thus large heat capacity in traditional superconductors, single-photon detector technologies in the ultralow photon energy regions (e.g., mid-infrared and terahertz frequencies) are extremely difficult to achieve. Notably, magic-angle TBG moiré superlattices can show dome-shaped superconducting states at a record-low electron density (~10^{12} cm^{-2}), and thus, an extremely small heat capacity (e.g., a few hundred times the Boltzmann constant) (Fig. 5H) (*133*, *134*). Consequently, a single photon, even with ultralow energy, can cause a sizable temperature increase to break down the superconducting state, offering extraordinary opportunities for single-photon detection in previously unattainable mid-infrared and even terahertz range (Fig. 5I) (*17*, *133*, *134*). Indeed, by quantifying the calorimetric photoresponses, it was demonstrated theoretically that magic-angle TBG can have an ultrabroad detection range of light at the single-photon level from the visible to sub-terahertz with an exceptionally fast response time of ~4 ns as well as a high energy resolution better than 1 THz (*134*). Further efforts are required to experimentally demonstrate such exciting single-photon devices.

In close resemblance to reconstructed photonics, moiré superlattices with lattice reconstruction would also lead to the renormalization of optoelectronic properties, such as the formation of photocurrent patterns (Fig. 5J) (*135*, *136*). In addition, by integrating moiré superlattices with other structures, such as waveguides, cavities, plasmonics, and ring resonators, a large portfolio of fascinating optoelectronic devices with superior performance would emerge, e.g., single photon light-emitting devices, optical modulators, photoinduced valley currents, etc.

Perspectives and conclusions

The field of moiré photonics and optoelectronics is progressing rapidly. A diverse series of moiré photonic and optoelectronic phenomena have been witnessed in the span of less than 5 years, including but not limited to moiré excitons and polaritons, resonantly hybridized excitons, reconstructed collective excitations, strong mid-infrared photoresponses, intelligent light sensors, and terahertz single-photon detection. This not only opens up exciting possibilities for a wide range of new frontiers in basic scientific research, but also outlines a bright vision for various sophisticated physical phenomena and emerging technological innovations. Moreover, the dizzying pace of recent achievements suggests that we have only seen the tip of the iceberg, and there will be many more surprises to come in moiré photonics and optoelectronics.

Advanced scanning probe techniques

One future direction is to develop new, advanced scanning probe techniques with an ultrahigh spatial resolution (e.g., <5 nm), and therefore that can probe the photonic and optoelectronic properties in an individual moiré supercell. Currently, the measurements of moiré photonics and optoelectronics are based on far-field techniques with a spot size of ~1 μm. As a consequence, the detected results are a collection of signals from more than 10,000 moiré unit cells. Owing to the ubiquity of twist-angle inhomogeneity and strain from the fabrication of devices, the observed results are often highly complex and lack reproducibility. For instance, some moiré photonic and optoelectronic phenomena have been reported by only a single group and are yet to be reproduced by others. Developing new techniques that can locally probe the photonic and optoelectronic properties of different high-symmetry points within a single moiré unit cell [such as 4D scanning TEM spectroscopy (*137*) and near-field scanning (*61*)] will undoubtedly advance our understanding of the current conundrums and contradictions and greatly facilitate the further development of moiré photonics and optoelectronics.

New moiré systems

Considering that there are thousands of 2D crystals and the current research on moiré photonics and optoelectronics is only focused on twisted graphene and twisted TMDCs, one of the particularly intriguing directions is to explore the exotic photonic and optoelectronic phenomena of new moiré systems, such as those moiré superlattices composed of 2D ferroelectric, magnetic, and multiferroic crystals. In particular, recent studies reveal that magnetic moiré superlattices can display a pattern of nanoscale domains that alternate with layered antiferromagnetic and ferromagnetic states (*41*, *42*). Consequently, magnetic moiré superlattices enable the control of magnetic order, symmetry breaking, and interlayer hybridization on the nanoscale (*138*), opening a huge swath of possibilities for engineering photonic and optoelectronic properties—for example, nonreciprocal second-harmonic generation (*139*), magnon-exciton coupling (*140*), and inelastic light scattering (*141*). Beyond the single moiré, a moiré landscape involving two or multiple single moiré superlattices provides another direction and has the potential to spark the next "gold rush" of new moiré photonics and optoelectronics. For instance, the long-sought boson exciton crystals have been recently evidenced in a $WSe_2/MoSe_2/WSe_2$ heterotrilayer consisting of two single moiré patterns, paving a way toward the exploration of quantum optoelectronics and quantum coherent phenomena (*32*). Further, moiré superlattices beyond 2D materials, such as 1D vdW moiré superlattices (*142*), and moiré artificial metamaterials (*143*, *144*) could also offer promising avenues for engineering moiré photonics and optoelectronics, e.g., localization-to-delocalization transition of light (*143*), magic-angle lasers (*144*), and optical solitons (*145*).

Engineering moiré photonics and optoelectronics

Another crucial direction is to control moiré photonic and optoelectronic properties through different external degrees of freedom (such as ultrafast optical excitations, electric and magnetic field, strain, twist angle, doping, pressure, and Floquet engineering). For instance, it has been theoretically elucidated that because of different local static dipole moments, the electric field can switch the global moiré potential minima from one high-symmetry point to another high-symmetry point via the quantum-confined Stark effect, enabling the possibilities of programming the spatial locations and optical selection rule of moiré-trapped excitons on demand (*66*). Experimentally, it has also been demonstrated that strain can play a role in tailoring the moiré potential landscape and optoelectronic response (*146*). To name a few, uniaxial strain can tune the array of quantum dot–like 0D traps into parallel stripes of 1D quantum wires, breaking the threefold rotational symmetry and changing the optical selection rule of moiré exciton from circular to linear (*146*). It was shown that the itinerant excitons excited by light can highly tune the spin-spin exchange interactions between moiré-trapped carriers and result in ferromagnetic order in WS_2/WSe_2 moiré superlattices, outlining the possibility of all-optical

control of emergent moiré photonic and optoelectronic phenomena (*120*).

Technical innovations

Particularly interesting fields where moiré optoelectronics have the potential to trigger new on-chip applications include, but are not limited to, high-performance moiré lasers, quantum simulation of the Bose-Hubbard model, and programmable quantum light sources. Recently, by integrating MoS_2/WSe_2 superlattices with a silicon topological nanocavity, ultralow-threshold broadband excitonic lasing at the technologically important telecommunication O-band (1260 to 1360 nm) has been achieved at room temperature (*147*). The moiré excitonic laser shows the highest spectral coherence of <0.1-nm linewidth among all 2D material–based laser systems studied so far, facilitating high-performance on-chip photonics and optoelectronics (e.g., electrically pumped wavelength-tunable moiré lasers) (*147*). Further, owing to enhanced Coulomb interactions, moiré superlattices offer opportunities to create a regular array of spatially ordered quantum emitters. In the case of moiré interlayer excitons with long lifetime and strong dipolar interactions, the perfect arrays provide a promising paradigm for the realization of quantum simulation of Bose-Hubbard model. It is noteworthy that the Bose-Hubbard model, which underlies rich bosonic states of matter and quantum phase transitions, is challenging to achieve in solid-state systems. Theoretically, a wide variety of bosonic quantum matter phases and quantum phase transition associated with quantum-dot–like interlayer excitons have been predicted in moiré superlattices, such as exciton supersolid, superfluid-insulator phase transition, and bosonic Mott insulator (*148*, *149*). Experimentally, exciton Mott insulator and exciton density waves have been recently reported in WSe_2/WS_2 and Coulomb-coupled WS_2/bilayer WSe_2/WS_2 dual moiré superlattices, respectively (*150*, *151*).

In addition, in the case of intralayer excitons with high quantum efficiency and relatively low sensitivity to disorder in the surrounding dielectric, the spatially ordered moiré exciton array could serve as a superbright quantum light source (*12*, *66*). In particular, moiré quantum light sources show high programmability with polarization configuration and wavelength controlled by optical pumping, electric field, and strain. Such programmable quantum light sources would provide a firm basis for the exploration of coherent quantum phenomena of optical bosons and the development of various quantum technologies (*70*), such as the generation of polarization-entangled photon pairs, Dicke superradiance, spin-photon interfaces, quantum information processing, encryption, and sensing. In addition, moiré superlattices, by integrating with other optical systems (e.g., optical cavities, resonances, silicon-based waveguides, and fibers), will undoubtedly outline a bright vision for numerous emerging photonic and optoelectronic applications, such as quantum nonlinear optics, lasers, imaging arrays, modulators, switches, and polarization devices.

REFERENCES AND NOTES

1. K. S. Novoselov, A. Mishchenko, A. Carvalho, A. H. Castro Neto, 2D materials and van der Waals heterostructures. *Science* **353**, aac9439 (2016). doi: 10.1126/science.aac9439; pmid: 27471306
2. C. Jin *et al.*, Ultrafast dynamics in van der Waals heterostructures. *Nat. Nanotechnol.* **13**, 994–1003 (2018). doi: 10.1038/s41565-018-0298-5; pmid: 30397296
3. L. Du *et al.*, Engineering symmetry breaking in 2D layered materials. *Nat. Rev. Phys.* **3**, 193–206 (2021). doi: 10.1038/s42254-020-00276-0
4. B. Huang *et al.*, Emergent phenomena and proximity effects in two-dimensional magnets and heterostructures. *Nat. Mater.* **19**, 1276–1289 (2020). doi: 10.1038/s41563-020-0791-8; pmid: 32948831
5. N. Briggs *et al.*, Atomically thin half-van der Waals metals enabled by confinement heteroepitaxy. *Nat. Mater.* **19**, 637–643 (2020). doi: 10.1038/s41563-020-0631-x; pmid: 32157191
6. Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, M. S. Strano, Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. *Nat. Nanotechnol.* **7**, 699–712 (2012). doi: 10.1038/nnano.2012.193; pmid: 23132225
7. L. Du *et al.*, 2D proximate quantum spin liquid state in atomic-thin α-RuCl3. *2D Mater.* **6**, 015014 (2018). doi: 10.1088/2053-1583/aaee29
8. A. K. Geim, I. V. Grigorieva, Van der Waals heterostructures. *Nature* **499**, 419–425 (2013). doi: 10.1038/nature12385; pmid: 23887427
9. Y. Liu *et al.*, Van der Waals heterostructures and devices. *Nat. Rev. Mater.* **1**, 16042 (2016). doi: 10.1038/natrevmats.2016.42
10. E. Y. Andrei *et al.*, The marvels of moiré materials. *Nat. Rev. Mater.* **6**, 201–206 (2021). doi: 10.1038/s41578-021-00284-1
11. Y. Xu *et al.*, Correlated insulating states at fractional fillings of moiré superlattices. *Nature* **587**, 214–218 (2020). doi: 10.1038/s41586-020-2868-6; pmid: 33177668
12. N. P. Wilson, W. Yao, J. Shan, X. Xu, Excitons and emergent quantum phenomena in stacked 2D semiconductors. *Nature* **599**, 383–392 (2021). doi: 10.1038/s41586-021-03979-1; pmid: 34789905
13. K. F. Mak, J. Shan, Semiconductor moiré materials. *Nat. Nanotechnol.* **17**, 686–695 (2022). doi: 10.1038/s41565-022-01165-6; pmid: 35836003
14. R. Bistritzer, A. H. MacDonald, Moire bands in twisted double-layer graphene. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12233–12237 (2011). doi: 10.1073/pnas.1108174108; pmid: 21730173
15. L. A. Ponomarenko *et al.*, Cloning of Dirac fermions in graphene superlattices. *Nature* **497**, 594–597 (2013). doi: 10.1038/nature12187; pmid: 23676678
16. C. R. Dean *et al.*, Hofstadter's butterfly and the fractal quantum Hall effect in moiré superlattices. *Nature* **497**, 598–602 (2013). doi: 10.1038/nature12186; pmid: 23676673
17. Y. Cao *et al.*, Unconventional superconductivity in magic-angle graphene superlattices. *Nature* **556**, 43–50 (2018). doi: 10.1038/nature26160; pmid: 29512651
18. Y. Cao *et al.*, Correlated insulator behaviour at half-filling in magic-angle graphene superlattices. *Nature* **556**, 80–84 (2018). doi: 10.1038/nature26154; pmid: 29512654
19. K. L. Seyler *et al.*, Signatures of moiré-trapped valley excitons in $MoSe_2/WSe_2$ heterobilayers. *Nature* **567**, 66–70 (2019). doi: 10.1038/s41586-019-0957-1; pmid: 30804526
20. C. Jin *et al.*, Observation of moiré excitons in WSe_2/WS_2 heterostructure superlattices. *Nature* **567**, 76–80 (2019). doi: 10.1038/s41586-019-0976-y; pmid: 30804525
21. K. Tran *et al.*, Evidence for moiré excitons in van der Waals heterostructures. *Nature* **567**, 71–75 (2019). doi: 10.1038/s41586-019-0975-z; pmid: 30804527
22. E. M. Alexeev *et al.*, Resonantly hybridized excitons in moiré superlattices in van der Waals heterostructures. *Nature* **567**, 81–86 (2019). doi: 10.1038/s41586-019-0986-9; pmid: 30842637
23. H. Baek *et al.*, Highly energy-tunable quantum light from moiré-trapped excitons. *Sci. Adv.* **6**, eaba8526 (2020). doi: 10.1126/sciadv.aba8526; pmid: 32917702
24. A. L. Sharpe *et al.*, Emergent ferromagnetism near three-quarters filling in twisted bilayer graphene. *Science* **365**, 605–608 (2019). doi: 10.1126/science.aaw3780; pmid: 31346139
25. G. Chen *et al.*, Tunable correlated Chern insulator and ferromagnetism in a moiré superlattice. *Nature* **579**, 56–61 (2020). doi: 10.1038/s41586-020-2049-7; pmid: 32132694
26. M. Serlin *et al.*, Intrinsic quantized anomalous Hall effect in a moiré heterostructure. *Science* **367**, 900–903 (2020). doi: 10.1126/science.aay5533; pmid: 31857492
27. E. C. Regan *et al.*, Mott and generalized Wigner crystal states in WSe_2/WS_2 moiré superlattices. *Nature* **579**, 359–363 (2020). doi: 10.1038/s41586-020-2092-4; pmid: 32188951
28. X. Huang *et al.*, Correlated insulating states at fractional fillings of the WS_2/WSe_2 moiré lattice. *Nat. Phys.* **17**, 715–719 (2021). doi: 10.1038/s41567-021-01171-w
29. Y. Jiang *et al.*, Charge order and broken rotational symmetry in magic-angle twisted bilayer graphene. *Nature* **573**, 91–95 (2019). doi: 10.1038/s41586-019-1460-4; pmid: 31365921
30. C. Jin *et al.*, Stripe phases in WSe_2/WS_2 moiré superlattices. *Nat. Mater.* **20**, 940–944 (2021). doi: 10.1038/s41563-021-00959-8; pmid: 33767398
31. M. Haavisto, J. Lado, A. O. Fumega, Topological multiferroic order in twisted transition metal dichalcogenide bilayers. *SciPost Phys.* **13**, 052 (2022). doi: 10.21468/SciPostPhys.13.3.052
32. Y. Bai *et al.*, Evidence for Exciton Crystals in a 2D Semiconductor Heterotrilayer. arXiv:2207.09601 (2022).
33. C. Ma *et al.*, Intelligent infrared sensing enabled by tunable moiré quantum geometry. *Nature* **604**, 266–272 (2022). doi: 10.1038/s41586-022-04548-w; pmid: 35418636
34. C. N. Lau, M. W. Bockrath, K. F. Mak, F. Zhang, Reproducibility in the fabrication and physics of moiré materials. *Nature* **602**, 41–50 (2022). doi: 10.1038/s41586-021-04173-z; pmid: 35110759
35. D. Huang, J. Choi, C.-K. Shih, X. Li, Excitons in semiconductor moiré superlattices. *Nat. Nanotechnol.* **17**, 227–238 (2022). doi: 10.1038/s41565-021-01068-y; pmid: 35288673
36. E. C. Regan *et al.*, Emerging exciton physics in transition metal dichalcogenide heterobilayers. *Nat. Rev. Mater.* **7**, 778–795 (2022). doi: 10.1038/s41578-022-00440-1
37. K. Kim *et al.*, van der Waals Heterostructures with High Accuracy Rotational Alignment. *Nano Lett.* **16**, 1989–1995 (2016). doi: 10.1021/acs.nanolett.5b05263; pmid: 26859527
38. R. Frisenda *et al.*, Recent progress in the assembly of nanodevices and van der Waals heterostructures by deterministic placement of 2D materials. *Chem. Soc. Rev.* **47**, 53–68 (2018). doi: 10.1039/C7CS00556C; pmid: 29111548
39. L. Wang *et al.*, Correlated electronic phases in twisted bilayer transition metal dichalcogenides. *Nat. Mater.* **19**, 861–866 (2020). doi: 10.1038/s41563-020-0708-6; pmid: 32572205
40. J. Sung *et al.*, Broken mirror symmetry in excitonic response of reconstructed domains in twisted $MoSe_2/MoSe_2$ bilayers. *Nat. Nanotechnol.* **15**, 750–754 (2020). doi: 10.1038/s41565-020-0728-z; pmid: 32661373
41. T. Song *et al.*, Direct visualization of magnetic domains and moiré magnetism in twisted 2D magnets. *Science* **374**, 1140–1144 (2021). doi: 10.1126/science.abj7478; pmid: 34822270
42. Y. Xu *et al.*, Coexisting ferromagnetic-antiferromagnetic state in twisted bilayer CrI_3. *Nat. Nanotechnol.* **17**, 143–147 (2022). doi: 10.1038/s41565-021-01014-y; pmid: 34845332
43. Y. Tang *et al.*, Simulation of Hubbard model physics in WSe_2/WS_2 moiré superlattices. *Nature* **579**, 353–358 (2020). doi: 10.1038/s41586-020-2085-3; pmid: 32188950
44. A. J. Mannix *et al.*, Robotic four-dimensional pixel assembly of van der Waals solids. *Nat. Nanotechnol.* **17**, 361–366 (2022). doi: 10.1038/s41565-021-01061-5; pmid: 35075299
45. S. Masubuchi *et al.*, Autonomous robotic searching and assembly of two-dimensional crystals to build van der Waals superlattices. *Nat. Commun.* **9**, 1413 (2018). doi: 10.1038/s41467-018-03723-w; pmid: 29650955
46. Q. Wang *et al.*, Layer-by-layer epitaxy of multi-layer MoS_2 wafers. *Natl. Sci. Rev.* **9**, nwac077 (2022). doi: 10.1093/nsr/nwac077; pmid: 35769232
47. P. Solís-Fernández, H. Ago, Machine Learning Determination of the Twist Angle of Bilayer Graphene by Raman Spectroscopy: Implications for van der Waals

Heterostructures. *ACS Appl. Nano Mater.* **5**, 1356–1366 (2022). doi: 10.1021/acsanm.1c03928
48. P. Chen, Z. Zhang, X. Duan, X. Duan, Chemical synthesis of two-dimensional atomic crystals, heterostructures and superlattices. *Chem. Soc. Rev.* **47**, 3129–3151 (2018). doi: 10.1039/C7CS00887B; pmid: 29528342
49. W. Yang *et al*., Epitaxial growth of single-domain graphene on hexagonal boron nitride. *Nat. Mater.* **12**, 792–797 (2013). doi: 10.1038/nmat3695; pmid: 23852399
50. L. Yuan *et al*., Twist-angle-dependent interlayer exciton diffusion in WS_2-WSe_2 heterobilayers. *Nat. Mater.* **19**, 617–623 (2020). doi: 10.1038/s41563-020-0670-3; pmid: 32393806
51. C. Zhang *et al*., Interlayer couplings, Moiré patterns, and 2D electronic superlattices in MoS_2/WSe_2 hetero-bilayers. *Sci. Adv.* **3**, e1601459 (2017). doi: 10.1126/sciadv.1601459; pmid: 28070558
52. L. Du *et al*., Modulating PL and electronic structures of MoS_2/graphene heterostructures via interlayer twisting angle. *Appl. Phys. Lett.* **111**, 263106 (2017). doi: 10.1063/1.5011120
53. R. Ribeiro-Palau *et al*., Twistable electronics with dynamically rotatable heterostructures. *Science* **361**, 690–693 (2018). doi: 10.1126/science.aat6981
54. M. Liao *et al*., Ultra-low friction and edge-pinning effect in large-lattice-mismatch van der Waals heterostructures. *Nat. Mater.* **21**, 47–53 (2022). doi: 10.1038/s41563-021-01058-4; pmid: 34354215
55. S. M. Akkanen, H. A. Fernandez, Z. Sun, Optical Modification of 2D Materials: Methods and Applications. *Adv. Mater.* **34**, e2110152 (2022). doi: 10.1002/adma.202110152; pmid: 35139583
56. D. Wang *et al*., Thermally Induced Graphene Rotation on Hexagonal Boron Nitride. *Phys. Rev. Lett.* **116**, 126101 (2016). doi: 10.1103/PhysRevLett.116.126101; pmid: 27058087
57. Z. Zhang *et al*., Flat bands in twisted bilayer transition metal dichalcogenides. *Nat. Phys.* **16**, 1093–1096 (2020). doi: 10.1038/s41567-020-0958-x
58. L. J. McGilly *et al*., Visualization of moiré superlattices. *Nat. Nanotechnol.* **15**, 580–584 (2020). doi: 10.1038/s41565-020-0708-3; pmid: 32572229
59. T. I. Andersen *et al*., Excitons in a reconstructed moiré potential in twisted WSe_2/WSe_2 homobilayers. *Nat. Mater.* **20**, 480–487 (2021). doi: 10.1038/s41563-020-00873-5; pmid: 33398121
60. A. Weston *et al*., Atomic reconstruction in twisted bilayers of transition metal dichalcogenides. *Nat. Nanotechnol.* **15**, 592–597 (2020). doi: 10.1038/s41565-020-0682-9; pmid: 32451502
61. S. S. Sunku *et al*., Photonic crystals for nano-light in moiré graphene superlattices. *Science* **362**, 1153–1156 (2018). doi: 10.1126/science.aau5144; pmid: 30523109
62. K. Lee *et al*., Ultrahigh-resolution scanning microwave impedance microscopy of moiré lattices and superstructures. *Sci. Adv.* **6**, eabd1919 (2020). doi: 10.1126/sciadv.abd1919; pmid: 33298449
63. J. Yu *et al*., Imaging Graphene Moiré Superlattices via Scanning Kelvin Probe Microscopy. *Nano Lett.* **21**, 3280–3286 (2021). doi: 10.1021/acs.nanolett.1c00609; pmid: 33749279
64. K.-Q. Lin *et al*., Large-Scale Mapping of Moiré Superlattices by Hyperspectral Raman Imaging. *Adv. Mater.* **33**, e2008333 (2021). doi: 10.1002/adma.202008333; pmid: 34242447
65. G. Wang *et al*., Colloquium: Excitons in atomically thin transition metal dichalcogenides. *Rev. Mod. Phys.* **90**, 021001 (2018). doi: 10.1103/RevModPhys.90.021001
66. H. Yu, G.-B. Liu, J. Tang, X. Xu, W. Yao, Moiré excitons: From programmable quantum emitter arrays to spin-orbit-coupled artificial lattices. *Sci. Adv.* **3**, e1701696 (2017). doi: 10.1126/sciadv.1701696; pmid: 29152568
67. F. Wu, T. Lovorn, A. H. MacDonald, Topological Exciton Bands in Moiré Heterojunctions. *Phys. Rev. Lett.* **118**, 147401 (2017). doi: 10.1103/PhysRevLett.118.147401; pmid: 28430504
68. M. Brotons-Gisbert *et al*., Spin-layer locking of interlayer excitons trapped in moiré potentials. *Nat. Mater.* **19**, 630–636 (2020). doi: 10.1038/s41563-020-0687-7; pmid: 32451512
69. M. Brotons-Gisbert *et al*., Moiré-Trapped Interlayer Trions in a Charge-Tunable $WSe_2/MoSe_2$ Heterobilayer. *Phys. Rev. X* **11**, 031033 (2021). doi: 10.1103/PhysRevX.11.031033
70. M. Turunen *et al*., Quantum photonics with layered 2D materials. *Nat. Rev. Phys.* **4**, 219–236 (2022). doi: 10.1038/s42254-021-00408-0
71. S. Zhao *et al*., Excitons in mesoscopically reconstructed moiré hetersotructures. *arXiv*:2202.11139 (2022).
72. J. Wang *et al*., Diffusivity Reveals Three Distinct Phases of Interlayer Excitons in $MoSe_2/WSe_2$ Heterobilayers. *Phys. Rev. Lett.* **126**, 106804 (2021). doi: 10.1103/PhysRevLett.126.106804; pmid: 33784140
73. E. Liu *et al*., Signatures of moiré trions in $WSe_2/MoSe_2$ heterobilayers. *Nature* **594**, 46–50 (2021). doi: 10.1038/s41586-021-03541-z; pmid: 34079140
74. H. Baek *et al*., Optical read-out of Coulomb staircases in a moiré superlattice via trapped interlayer trions. *Nat. Nanotechnol.* **16**, 1237–1243 (2021). doi: 10.1038/s41565-021-00970-9; pmid: 34556832
75. X. Wang *et al*., Moiré trions in $MoSe_2/WSe_2$ heterobilayers. *Nat. Nanotechnol.* **16**, 1208–1213 (2021). doi: 10.1038/s41565-021-00969-2; pmid: 34531556
76. K. F. Mak, D. Xiao, J. Shan, Light-valley interactions in 2D semiconductors. *Nat. Photonics* **12**, 451–460 (2018). doi: 10.1038/s41566-018-0204-6
77. H. Yu, Y. Wang, Q. Tong, X. Xu, W. Yao, Anomalous Light Cones and Valley Optical Selection Rules of Interlayer Excitons in Twisted Heterobilayers. *Phys. Rev. Lett.* **115**, 187002 (2015). doi: 10.1103/PhysRevLett.115.187002; pmid: 26565491
78. S. Brem, C. Linderälv, P. Erhart, E. Malic, Tunable Phases of Moiré Excitons in van der Waals Heterostructures. *Nano Lett.* **20**, 8534–8540 (2020). doi: 10.1021/acs.nanolett.0c03019; pmid: 32970445
79. F. Wu, T. Lovorn, A. H. MacDonald, Theory of optical absorption by interlayer excitons in transition metal dichalcogenide heterobilayers. *Phys. Rev. B* **97**, 035306 (2018). doi: 10.1103/PhysRevB.97.035306
80. T. Smoleński *et al*., Signatures of Wigner crystal of electrons in a monolayer semiconductor. *Nature* **595**, 53–57 (2021). doi: 10.1038/s41586-021-03590-4; pmid: 34194018
81. N. Zhang *et al*., Moiré Intralayer Excitons in a $MoSe_2/MoS_2$ Heterostructure. *Nano Lett.* **18**, 7651–7657 (2018). doi: 10.1021/acs.nanolett.8b03266; pmid: 30403876
82. M. H. Naik *et al*., Intralayer charge-transfer moiré excitons in van der Waals superlattices. *Nature* **609**, 52–57 (2022). doi: 10.1038/s41586-022-04991-9; pmid: 36045239
83. A. Ciarrocchi *et al*., Polarization switching and electrical control of interlayer excitons in two-dimensional van der Waals heterostructures. *Nat. Photonics* **13**, 131–136 (2019). doi: 10.1038/s41566-018-0325-y; pmid: 30886643
84. E. Marcellina *et al*., Evidence for Moiré Trions in Twisted $MoSe_2$ Homobilayers. *Nano Lett.* **21**, 4461–4468 (2021). doi: 10.1021/acs.nanolett.1c01207; pmid: 33970625
85. A. Righi *et al*., Graphene moiré patterns observed by umklapp double-resonance Raman scattering. *Phys. Rev. B Condens. Matter Mater. Phys.* **84**, 241409 (2011). doi: 10.1103/PhysRevB.84.241409
86. G. S. N. Eliel *et al*., Intralayer and interlayer electron-phonon interactions in twisted graphene heterostructures. *Nat. Commun.* **9**, 1221 (2018). doi: 10.1038/s41467-018-03479-3; pmid: 29572537
87. A. Eckmann *et al*., Raman fingerprint of aligned graphene/h-BN superlattices. *Nano Lett.* **13**, 5242–5246 (2013). doi: 10.1021/nl402679b; pmid: 24156357
88. M. Liao *et al*., Precise control of the interlayer twist angle in large scale MoS_2 homostructures. *Nat. Commun.* **11**, 2153 (2020). doi: 10.1038/s41467-020-16056-4; pmid: 32358571
89. D. A. Ruiz-Tijerina, V. I. Fal'ko, Interlayer hybridization and moiré superlattice minibands for electrons and excitons in heterobilayers of transition-metal dichalcogenides. *Phys. Rev. B* **99**, 125424 (2019). doi: 10.1103/PhysRevB.99.125424
90. L. Zhang *et al*., Twist-angle dependence of moiré excitons in $WS_2/MoSe_2$ heterobilayers. *Nat. Commun.* **11**, 5888 (2020). doi: 10.1038/s41467-020-19466-6; pmid: 33208738
91. Y. Tang *et al*., Tuning layer-hybridized moiré excitons by the quantum-confined Stark effect. *Nat. Nanotechnol.* **16**, 52–57 (2021). doi: 10.1038/s41565-020-00783-2; pmid: 33139934
92. Y. Shimazaki *et al*., Strongly correlated electrons and hybrid excitons in a moiré heterostructure. *Nature* **580**, 472–477 (2020). doi: 10.1038/s41586-020-2191-2; pmid: 32322064
93. H. Yu, W. Yao, Luminescence Anomaly of Dipolar Valley Excitons in Homobilayer Semiconductor Moiré Superlattices. *Phys. Rev. X* **11**, 021042 (2021). doi: 10.1103/PhysRevX.11.021042
94. S. Brem *et al*., Hybridized intervalley moiré excitons and flat bands in twisted WSe_2 bilayers. *Nanoscale* **12**, 11088–11094 (2020). doi: 10.1039/D0NR02160A; pmid: 32400821
95. C. Zhang *et al*., Systematic study of electronic structure and band alignment of monolayer transition metal dichalcogenides in Van der Waals heterostructures. *2D Mater.* **4**, 015026 (2016). doi: 10.1088/2053-1583/4/1/015026
96. P. Rivera *et al*., Interlayer valley excitons in heterobilayers of transition metal dichalcogenides. *Nat. Nanotechnol.* **13**, 1004–1015 (2018). doi: 10.1038/s41565-018-0193-0; pmid: 30104622
97. Y. Zhao *et al*., Interlayer exciton complexes in bilayer MoS_2. *Phys. Rev. B* **105**, L041411 (2022). doi: 10.1103/PhysRevB.105.L041411
98. C. Schneider, M. M. Glazov, T. Korn, S. Höfling, B. Urbaszek, Two-dimensional semiconductors in the regime of strong light-matter coupling. *Nat. Commun.* **9**, 2695 (2018). doi: 10.1038/s41467-018-04866-6; pmid: 30002368
99. D. N. Basov, M. M. Fogler, F. J. García de Abajo, Polaritons in van der Waals materials. *Science* **354**, aag1992 (2016). doi: 10.1126/science.aag1992; pmid: 27738142
100. T. Low *et al*., Polaritons in layered two-dimensional materials. *Nat. Mater.* **16**, 182–194 (2017). doi: 10.1038/nmat4792; pmid: 27893724
101. L. Zhang *et al*., Van der Waals heterostructure polaritons with moiré-induced nonlinearity. *Nature* **591**, 61–65 (2021). doi: 10.1038/s41586-021-03228-5; pmid: 33658695
102. G. X. Ni *et al*., Plasmons in graphene moiré superlattices. *Nat. Mater.* **14**, 1217–1222 (2015). doi: 10.1038/nmat4425; pmid: 26413987
103. N. C. H. Hesp *et al*., Observation of interband collective excitations in twisted bilayer graphene. *Nat. Phys.* **17**, 1162–1168 (2021). doi: 10.1038/s41567-021-01327-8
104. T. Huang *et al*., Observation of chiral and slow plasmons in twisted bilayer graphene. *Nature* **605**, 63–68 (2022). doi: 10.1038/s41586-022-04520-8; pmid: 35508778
105. C. Lewandowski, L. Levitov, Intrinsically undamped plasmon modes in narrow electron bands. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20869–20874 (2019). doi: 10.1073/pnas.1909069116; pmid: 31562197
106. T. Stauber, H. Kohler, Quasi-Flat Plasmonic Bands in Twisted Bilayer Graphene. *Nano Lett.* **16**, 6844–6849 (2016). doi: 10.1021/acs.nanolett.6b02587; pmid: 27776411
107. L. Brey, T. Stauber, T. Slipchenko, L. Martín-Moreno, Plasmonic Dirac Cone in Twisted Bilayer Graphene. *Phys. Rev. Lett.* **125**, 256804 (2020). doi: 10.1103/PhysRevLett.125.256804; pmid: 33416378
108. G. Hu *et al*., Topological polaritons and photonic magic angles in twisted α-MoO_3 bilayers. *Nature* **582**, 209–213 (2020). doi: 10.1038/s41586-020-2359-9; pmid: 32528096
109. M. Chen *et al*., Configurable phonon polaritons in twisted α-MoO_3. *Nat. Mater.* **19**, 1307–1311 (2020). doi: 10.1038/s41563-020-0732-6; pmid: 32661384
110. Z. Zhang *et al*., Correlated interlayer exciton insulator in heterostructures of monolayer WSe_2 and moiré WS_2/WSe_2. *Nat. Phys.* **18**, 1214–1220 (2022). doi: 10.1038/s41567-022-01702-z
111. J. Gu *et al*., Dipolar excitonic insulator in a moiré lattice. *Nat. Phys.* **18**, 395–400 (2022). doi: 10.1038/s41567-022-01532-z
112. H. Li *et al*., Imaging two-dimensional generalized Wigner crystals. *Nature* **597**, 650–654 (2021). doi: 10.1038/s41586-021-03874-9; pmid: 34588665
113. L. Ma *et al*., Strongly correlated excitonic insulator in atomic double layers. *Nature* **598**, 585–589 (2021). doi: 10.1038/s41586-021-03947-9; pmid: 34707306
114. E. Liu *et al*., Excitonic and Valley-Polarization Signatures of Fractional Correlated Electronic Phases in a WSe_2/WS_2 Moiré Superlattice. *Phys. Rev. Lett.* **127**, 037402 (2021). doi: 10.1103/PhysRevLett.127.037402; pmid: 34328773
115. S. Miao *et al*., Strong interaction between interlayer excitons and correlated electrons in WSe_2/WS_2 moiré superlattice. *Nat. Commun.* **12**, 3608 (2021). doi: 10.1038/s41467-021-23732-6; pmid: 34127668
116. Y. Shimazaki *et al*., Optical Signatures of Periodic Charge Distribution in a Mott-like Correlated Insulator State. *Phys. Rev. X* **11**, 021027 (2021). doi: 10.1103/PhysRevX.11.021027
117. L. Du *et al*., Giant anisotropic photonics in the 1D van der Waals semiconductor fibrous red phosphorus. *Nat. Commun.* **12**, 4822 (2021). doi: 10.1038/s41467-021-25104-6; pmid: 34376660
118. A. J. Campbell *et al*., Exciton-polarons in the presence of strongly correlated electronic states in a $MoSe_2/WSe_2$ moiré superlattice. *npj 2D Mater. Appl.* **6**, 79 (2022). doi: 10.1038/s41699-022-00358-w
119. J. Liu, X. Dai, Anomalous Hall effect, magneto-optical properties, and nonlinear optical properties of twisted graphene systems. *npj Computat. Mater.* **6**, 57 (2020). doi: 10.3390/ma14010057

120. X. Wang *et al*., Light-induced ferromagnetism in moiré superlattices. *Nature* **604**, 468–473 (2022). doi: 10.1038/s41586-022-04472-z; pmid: 35444320
121. H. Yoo *et al*., Atomic and electronic reconstruction at the van der Waals interface in twisted bilayer graphene. *Nat. Mater.* **18**, 448–453 (2019). doi: 10.1038/s41563-019-0346-z; pmid: 30988451
122. V. V. Enaldiev, V. Zólyomi, C. Yelgel, S. J. Magorrian, V. I. Fal'ko, Stacking Domains and Dislocation Networks in Marginally Twisted Bilayers of Transition Metal Dichalcogenides. *Phys. Rev. Lett.* **124**, 206101 (2020). doi: 10.1103/PhysRevLett.124.206101; pmid: 32501062
123. L. Jiang *et al*., Soliton-dependent plasmon reflection at bilayer graphene domain walls. *Nat. Mater.* **15**, 840–844 (2016). doi: 10.1038/nmat4653; pmid: 27240109
124. J. Quan *et al*., Phonon renormalization in reconstructed MoS_2 moiré superlattices. *Nat. Mater.* **20**, 1100–1105 (2021). doi: 10.1038/s41563-021-00960-1; pmid: 33753933
125. A. C. Gadelha *et al*., Localization of lattice dynamics in low-angle twisted bilayer graphene. *Nature* **590**, 405–409 (2021). doi: 10.1038/s41586-021-03252-5; pmid: 33597759
126. B. Deng *et al*., Strong mid-infrared photoresponse in small-twist-angle bilayer graphene. *Nat. Photonics* **14**, 549–553 (2020). doi: 10.1038/s41566-020-0644-7
127. J. Yang *et al*., Spectroscopy signatures of electron correlations in a trilayer graphene/hBN moiré superlattice. *Science* **375**, 1295–1299 (2022). doi: 10.1126/science.abg3036; pmid: 35298267
128. G. Chen *et al*., Signatures of tunable superconductivity in a trilayer graphene moiré superlattice. *Nature* **572**, 215–219 (2019). doi: 10.1038/s41586-019-1393-y pmid: 31316203
129. Q. Ma, A. G. Grushin, K. S. Burch, Topology and geometry under the nonlinear electromagnetic spotlight. *Nat. Mater.* **20**, 1601–1614 (2021). doi: 10.1038/s41563-021-00992-7; pmid: 34127824
130. M. Otteneder *et al*., Terahertz Photogalvanics in Twisted Bilayer Graphene Close to the Second Magic Angle. *Nano Lett.* **20**, 7152–7158 (2020). doi: 10.1021/acs.nanolett.0c02474; pmid: 32915581
131. P. He *et al*., Graphene moiré superlattices with giant quantum nonlinearity of chiral Bloch electrons. *Nat. Nanotechnol.* **17**, 378–383 (2022). doi: 10.1038/s41565-021-01060-6; pmid: 35115723
132. M. Huang *et al*., Giant nonlinear Hall effect in twisted bilayer WSe_2. *Natl. Sci. Rev.* 10.1093/nsr/nwac232 (2022). doi: 10.1093/nsr/nwac232
133. X. Lu *et al*., Superconductors, orbital magnets and correlated states in magic-angle bilayer graphene. *Nature* **574**, 653–657 (2019). doi: 10.1038/s41586-019-1695-0; pmid: 31666722
134. P. Seifert *et al*., Magic-Angle Bilayer Graphene Nanocalorimeters: Toward Broadband, Energy-Resolving Single Photon Detection. *Nano Lett.* **20**, 3459–3464 (2020). doi: 10.1021/acs.nanolett.0c00373; pmid: 32315186
135. S. S. Sunku *et al*., Hyperbolic enhancement of photocurrent patterns in minimally twisted bilayer graphene. *Nat. Commun.* **12**, 1641 (2021). doi: 10.1038/s41467-021-21792-2; pmid: 33712611
136. N. C. H. Hesp *et al*., Nano-imaging photoresponse in a moiré unit cell of minimally twisted bilayer graphene. *Nat. Commun.* **12**, 1640 (2021). doi: 10.1038/s41467-021-21862-5; pmid: 33712606
137. X. Yan *et al*., Single-defect phonons imaged by electron microscopy. *Nature* **589**, 65–69 (2021). doi: 10.1038/s41586-020-03049-y; pmid: 33408374
138. N. P. Wilson *et al*., Interlayer electronic coupling on demand in a 2D magnetic semiconductor. *Nat. Mater.* **20**, 1657–1662 (2021). doi: 10.1038/s41563-021-01070-8; pmid: 34312534
139. Z. Sun *et al*., Giant nonreciprocal second-harmonic generation from antiferromagnetic bilayer CrI_3. *Nature* **572**, 497–501 (2019). doi: 10.1038/s41586-019-1445-3; pmid: 31367036
140. Y. J. Bae *et al*., Exciton-coupled coherent magnons in a 2D semiconductor. *Nature* **609**, 282–286 (2022). doi: 10.1038/s41586-022-05024-1; pmid: 36071189
141. B. Huang *et al*., Tuning inelastic light scattering via symmetry control in the two-dimensional magnet CrI_3. *Nat. Nanotechnol.* **15**, 212–216 (2020). doi: 10.1038/s41565-019-0598-4; pmid: 31907441
142. S. Zhao, R. Kitaura, P. Moon, M. Koshino, F. Wang, Interlayer Interactions in 1D Van der Waals Moiré Superlattices. *Adv. Sci. (Weinh.)* **9**, e2103460 (2022). doi: 10.1002/advs.202103460; pmid: 34841726
143. P. Wang *et al*., Localization and delocalization of light in photonic moiré lattices. *Nature* **577**, 42–46 (2020). doi: 10.1038/s41586-019-1851-6; pmid: 31853062
144. X.-R. Mao, Z.-K. Shao, H.-Y. Luan, S.-L. Wang, R.-M. Ma, Magic-angle lasers in nanostructured moiré superlattice. *Nat. Nanotechnol.* **16**, 1099–1105 (2021). doi: 10.1038/s41565-021-00956-7; pmid: 34400821
145. Q. Fu *et al*., Optical soliton formation controlled by angle twisting in photonic moiré lattices. *Nat. Photonics* **14**, 663–668 (2020). doi: 10.1038/s41566-020-0679-9
146. Y. Bai *et al*., Excitons in strain-induced one-dimensional moiré potentials at transition metal dichalcogenide heterojunctions. *Nat. Mater.* **19**, 1068–1073 (2020). doi: 10.1038/s41563-020-0730-8; pmid: 32661380
147. Q. Lin *et al*., A room-temperature moiré interlayer exciton laser. *arXiv:2302.01266* (2023).
148. Y.-H. Zhang, D. N. Sheng, A. Vishwanath, SU (4) Chiral Spin Liquid, Exciton Supersolid, and Electric Detection in Moiré Bilayers. *Phys. Rev. Lett.* **127**, 247701 (2021). doi: 10.1103/PhysRevLett.127.247701; pmid: 34951785
149. Z. Bi, L. Fu, Excitonic density wave and spin-valley superfluid in bilayer transition metal dichalcogenide. *Nat. Commun.* **12**, 642 (2021). doi: 10.1038/s41467-020-20802-z; pmid: 33510138
150. R. Xiong *et al*., Bosonic Mott insulator in WSe_2/WS_2 moiré superlattice. arXiv:2207.10764 [cond-mat.str-el] (2022).
151. Y. Zeng *et al*., Exciton density waves in Coulomb-coupled dual moiré lattices. *Nat. Mater.* **22**, 175–179 (2023). doi: 10.1038/s41563-022-01454-4; pmid: 36635591

ACKNOWLEDGMENTS

We acknowledge numerous colleagues with whom we have had the opportunity to work and gain knowledge in the constantly growing field of moiré photonics and optoelectronics. Z.S. also thanks J. Bai and Z. Ren for discussion and assistance during his current visiting professorship at the Northwestern University, China. **Funding:** The authors gratefully acknowledge the financial support from the National Science Foundation of China (NSFC) (grant nos. 12274447, 61888102, and 11834017), the National Key Research and Development Program (grant no. 2021YFA1202900), the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDB30000000), and the Key-Area Research and Development Program of Guangdong Province (grant no. 2020B0101340001), the Academy of Finland (314810, 333982, 336144, 333099, 336818, 352780, 352930, and 353364), the Academy of Finland Flagship Programme (320167, PREIN), ERC (834742), the EU H2020-MSCA-RISE-872049 (IPN-Bio), the European Union's Horizon 2020 research and innovation program (965124, FEMTOCHIP), and the National Science Centre, Poland (grant no. 2018/31/B/ST3/02111). F.W. acknowledges the support from US Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under contract no. DE-AC02-05-CH11231 (van der Waals heterostructure program KCFW16). **Author contributions:** L.D., G.Z, and Z.S. conceived the idea and led the writing. M.M., Z.H., and F.W. contributed substantially to the discussion of the content. All authors reviewed and edited the manuscript before submission. **Competing interests:** The authors declare no competing interests.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

Submitted 28 November 2022; accepted 24 February 2023
10.1126/science.adg0014

WORKING LIFE

By **Andrea del Valle**

Finding my way home

I sat down for the interview from my office in Taiwan, where I'd moved from Guatemala to pursue my dream of becoming a scientist. My words would be used for a video produced by the Guatemalan government, and I hoped they would inspire young Guatemalans to follow my career path. I spoke about the joys of being a Ph.D. student, the pleasure of discovering new things every day, and how I felt I was in my natural element. There was one thing I didn't mention: my fear that my career as a researcher might prevent me from ever moving back home. But the video itself led to a solution. It helped me connect with the people and culture of Latin America—even if I was doing so from afar.

My scientific spark came when I was 14 years old. My biology teacher invited me to attend Guatemala's "scientific Olympics," which brought together students from across the country to participate in scientific activities. It was there that I first learned about DNA. I was fascinated by mutations and their links with many diseases, including cancer. Ever since then, all I have wanted to do is learn about cancer.



"I hoped [my words] would inspire young Guatemalans to follow my career path."

No institutions were researching cancer in Guatemala, so when the time came to think about universities, I looked abroad, eventually deciding to study in Taiwan. I planned to earn a bachelor's degree and then return home. At the time, a bachelor's was enough to secure a faculty position in Guatemala—in part because the positions often revolved around teaching, with very little research. I thought if I worked hard, perhaps I could do that, too.

But after an internship in a cancer research lab, which opened my eyes to the wonder of scientific discovery, I realized I didn't want to spend my career only teaching. I also wanted to do research.

That was a sobering realization because it meant I might have to stay abroad indefinitely. In Taiwan, a single cancer lab might spend thousands of dollars per week on research. In Guatemala, professors had very little research funding and well-equipped labs were almost nonexistent.

I decided to pursue a Ph.D. anyway, staying in Taiwan because I won an award that would pay for my graduate education. It was difficult to be so far away from my family and the culture I grew up in. But I enjoyed my work and felt fortunate to have the opportunity to spend my days researching cancer. I also found an unexpected way to connect to my home country, thanks to the video.

Shortly after it was posted online, I was inundated with messages and postcards from people in Guatemala. The outpouring of interest inspired me to travel back home and spend several weeks delivering talks at schools. One talk was to a group of girls who had been rescued from human trafficking. I showed them pictures of my laboratory, described my research, and had them do a small hands-on activity. Afterward, my cousin, who is a psychologist at the school, told me that some of the students expressed interest in becoming a scientist. That helped me see the impact I could have outside of the laboratory—and that even if I couldn't live and work in Guatemala, perhaps I could make a difference through outreach.

After returning to Taiwan, I started a Facebook page to communicate science to people in my home country. During the pandemic, I translated Chinese- or English-language news stories about COVID-19 into Spanish and posted them there. My audience grew to 5000 people in less than a month, and I started to get invitations to deliver virtual talks in Mexico, Peru, and Bolivia. During one talk, more than 10,000 viewers tuned in.

This isn't what I imagined I'd be doing when I left home to pursue a career as a scientist. And I'm still hopeful that someday I'll be able to move back to Guatemala and run a research lab as a professor there. But for now, I am thankful that I discovered how to make a difference in my own small way and transmit my passion for science to the next generation in Latin America. ■

Andrea del Valle is a postdoctoral researcher at the Karolinska Institute.

ILLUSTRATION: ROBERT NEUBECKER